Check for updates

Replicated radiation of a plant clade along a cloud forest archipelago

Michael J. Donoghue [1,2] ✉, Deren A. R. Eaton [3] ✉, Carlos A. Maya-Lastra [3], Michael J. Landis[4], Patrick W. Sweeney[2], Mark E. Olson [5], N. Ivalú Cacho [5], Morgan K. Moeglein [1], Jordan R. Gardner[1], Nora M. Heaphy [1], Matiss Castorena [5,6], Alí Segovia Rivas [5], Wendy L. Clement [7] and Erika J. Edwards [1,2] ✉
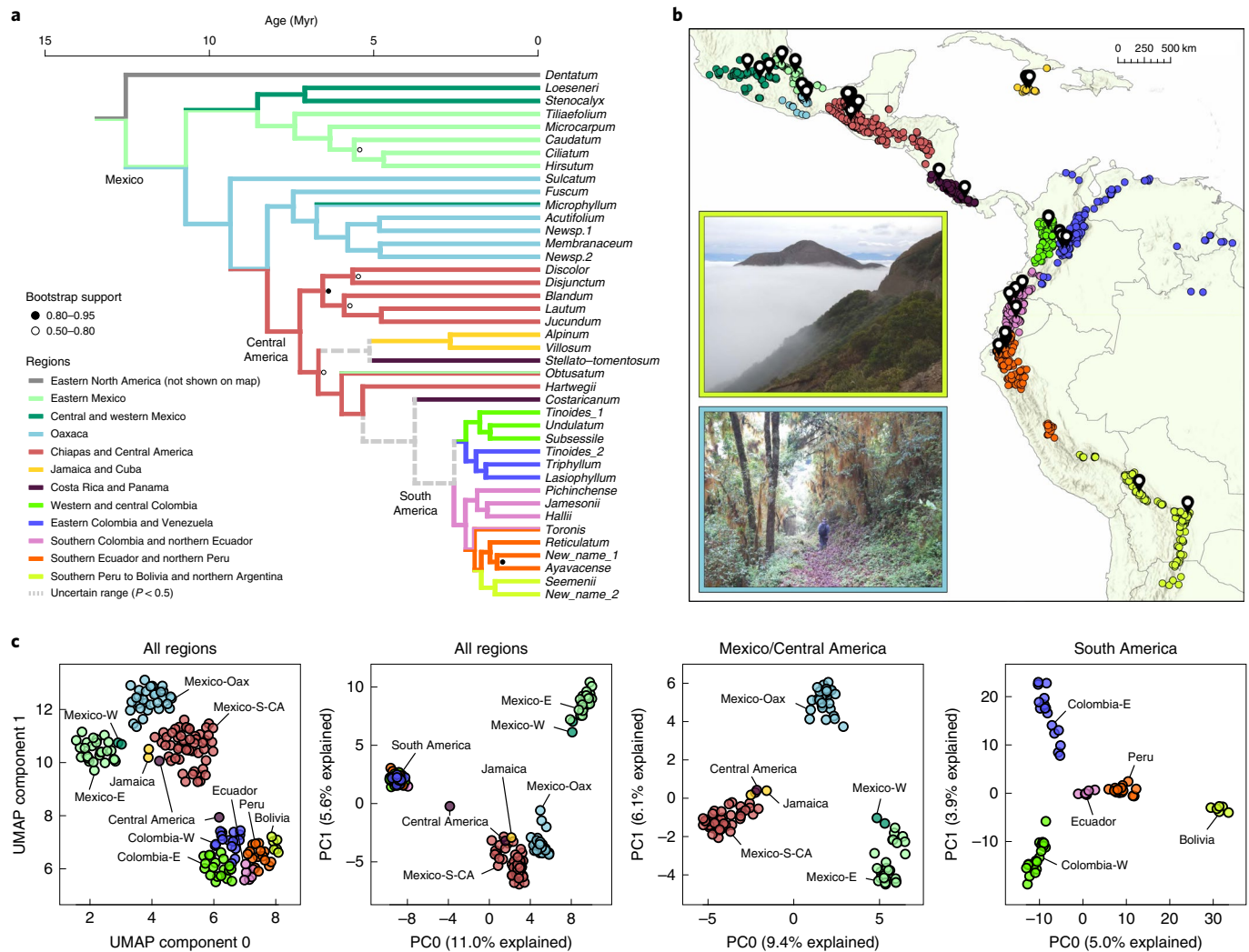
Replicated radiations, in which sets of similar forms evolve repeatedly within different regions, can provide powerful insights into parallel evolution and the assembly of functional diversity within communities. Several cases have been described in animals, but in plants we lack well-documented cases of replicated radiation that combine comprehensive phylogenetic and biogeographic analyses, the delimitation of geographic areas within which a set of 'ecomorphs' evolved independently and the identification of potential underlying mechanisms. Here we document the repeated evolution of a set of leaf ecomorphs in a group of neotropical plants. The *Oreinotinus* lineage within the angiosperm clade *Viburnum* spread from Mexico to Argentina through disjunct cloud forest environments. In 9 of 11 areas of endemism, species with similar sets of leaf forms evolved in parallel. We reject gene-flow-mediated evolution of similar leaves and show, instead, that species with disparate leaf forms differ in their climatic niches, supporting ecological adaptation as the driver of parallelism. Our identification of a case of replicated radiation in plants sets the stage for comparative analyses of such phenomena across the tree of life.

Replicated radiations[1,2] have been documented in a number of animal groups, including *Anolis* lizards in the Caribbean[3,4], cichlid fishes in African rift lakes[5,6] and spiders in Hawaii[7,8]. In plants, case studies abound of convergent and parallel evolution, including the repeated evolution of specific modes of pollination (for example, ref. [9]), dispersal (for example, ref. [10]) and photosynthesis (for example, ref. [11]). Likewise, many studies have convincingly documented the rapid adaptive radiation and morphological diversification of plant lineages in connection with their geographic movements (for example, refs. [12–14]). These studies have not, however, provided a compelling example of replicated radiation[15]. While several publications have referred to this phenomenon in plants[16–21], none of these cases compares closely in scope, scale or certitude to the several now-classic examples in animals[3–8]. A key missing element is the delimitation of separate areas of endemism within which speciation and adaptation have resulted in the repeated evolution of a set of specific 'ecomorphs' (different forms occupying different ecological niches)[1,2]. Accordingly, a comparable botanical platform for in-depth analyses of parallel functional diversification is lacking, making it difficult to compare and contrast such radiations broadly across the tree of life[15]. Here we document the repeated evolution of a set of leaf ecomorphs in a clade of neotropical plants that has diversified through disjunct cloud forest environments whose geography resembles that of an island archipelago.

Viburnum (Adoxaceae, Dipsacales) includes ~165 species of shrubs and small trees with small white flowers pollinated by insects and fleshy single-seeded fruits dispersed by birds[22,23]. Two lineages of *Viburnum* entered North America from Asia ~40 Myr

ago (Ma) in the Eocene epoch[24]. Within one of these major lineages, the *Oreinotinus* clade[25] diversified rapidly[26] as it spread southwards from Mexico starting in the late Miocene epoch (~8–12 Ma). This radiation was triggered neither by a morphological innovation[25,26] nor a novel change in chromosome number or genome size[27]. Instead, it appears to have been promoted by a combination of factors[28,29], including a shift into patchily distributed cloud forest environments (characterized by frequent cloud cover at canopy level and limited seasonality, but large daily fluctuation in temperature[30,31]) and the evolution of an evergreen habit[32]. Today, ~42 *Oreinotinus* species can be found mainly between 1,000 and 3,000 m throughout the neotropics, with a number of species present in each of 11 mountainous areas of endemism that we have identified for the clade (Fig. 1a,b, Methods, Extended Data Fig. 3, Supplementary Tables 1 and 2 and Supplementary Fig. 2). Most of these areas are separated by well-known lowland barriers to dispersal for plants of high-elevation mesic forests, such as the Isthmus of Tehuantepec, the Isthmus of Panama and, in the case of Jamaica and Cuba, the Caribbean Sea. They also correspond to biogeographic provinces recognized in studies of other cloud forest organisms (for example, our southern Ecuador and northern Peru region corresponds to the Huancabamba Depression, and our eastern Colombia-Venezuela region corresponds to the Cordillera Oriental of ref. [30]; for Mexico see ref. [31]). Of the 40 *Oreinotinus* species studied here, only 4 are present in more than 1 of the 11 areas. This distribution pattern, combined with striking differences among these species in a set of leaf traits linked to plant function, provides an ideal system in which to explore the possibility of replicated radiation.

[1]Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT, USA. [2]Peabody Museum of Natural History, Yale University, New Haven, CT, USA. [3]Department of Ecology, Evolution & Environmental Biology, Columbia University, New York, NY, USA. [4]Department of Biology, Washington University in St. Louis, St. Louis, MO, USA. [5]Departamento de Botánica, Instituto de Biología, Universidad Nacional Autónoma de México, Ciudad de Mexico, Mexico. [6]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ, USA. [7]Department of Biology, The College of New Jersey, Ewing, NJ, USA. ✉e-mail: michael.donoghue@yale.edu; de2356@columbia.edu; erika.edwards@yale.edu
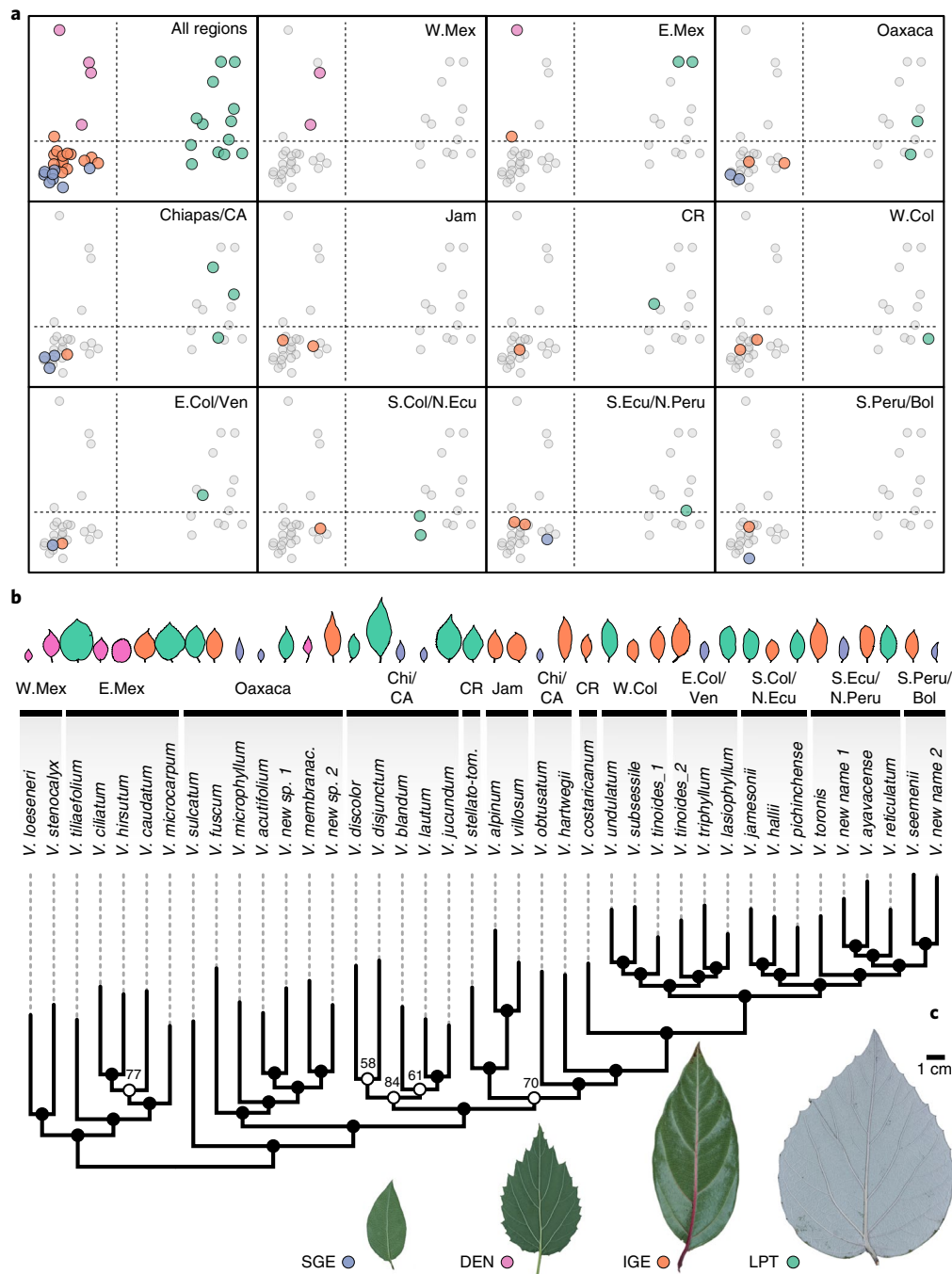
Fig. 1 | Phylogeny and biogeography of *Oreinotinus*. a, Historical biogeographic reconstruction for *Oreinotinus* inferred using a time-calibrated species-level phylogeny and a DEC model (Methods). Species-level tips are coloured by areas of endemism, and branch/node colours display the most probable (*P* > 0.5) ancestral ranges. Widespread lineages are assigned multiple colours, with one colour per inhabited area in the range. Dashed grey lines indicate that no one range was supported at *P* > 0.5. Phylogenetic relationships are well supported (bootstrap values of >95%) except as noted, and our inferences are robust to phylogenetic uncertainty marked at several nodes (Methods and Supplementary Fig. 3). **b**, The distribution of *Oreinotinus* in 11 areas of endemism in montane forests of the neotropics; coloured points mark herbarium specimen localities; pins mark locations of accessions in **a** (Methods and Extended Data Fig. 3). Photographs show clouds shrouding Tucumano forest in the vicinity of Tarija, Bolivia (top) and moss-covered understory of a cloud forest in the Santiago Comaltepec highlands, Oaxaca, Mexico (bottom). **c**, Dimensionality reduction methods applied to RAD-seq SNPs (Methods and Extended Data Fig. 4) show genetic similarities within and among areas of endemism. A UMAP projection distinguishes local and global structure matching geography; PCAs show differences among samples spanning the entire *Oreinotinus* range: Mexico, Central America and the Caribbean and South America. E=East, S=South, W=West, CA=Central America, Oax=Oaxaca.

Results and discussion

Phylogeny and biogeography. Our phylogenetic analyses are based on population genomic restriction site-associated DNA sequencing (RAD-seq) data for 180 individuals spanning the geographic range of *Oreinotinus*, mapped to a new draft reference genome assembly for *Viburnum lautum* (Methods and Supplementary Table 3). Analysis of the full dataset (Methods, Extended Data Fig. 1 and Supplementary Table 3) shows that accessions from multiple populations of the same morphology-based species form clades in all but five instances (discussed further in Supplementary Note 1). In two cases (involving the circumscription of *V. sulcatum* and *V. fuscum*) we will elsewhere describe new species to reflect our molecular and morphological findings. Specimens identified here as *Viburnum* 'New Name 1' appear to be paraphyletic with respect to

V. reticulatum plus *V. ayavacense*, but the support for these relationships is too weak to warrant a nomenclatural change. Bolivia has only recently been occupied by *Viburnum*, and we interpret the lack of reciprocal monophyly in this area as evidence of incomplete lineage sorting due to ongoing speciation. In the final case, *V. tinoides*, our accessions appear in two different clades. As we are currently unable to differentiate these populations based on morphology or ecology, *V. tinoides* is retained here, pending further study, as a single widespread species.

Our biogeographic analysis, conducted in a dispersal-extinction-cladogenesis (DEC) framework[33] and using a time-calibrated species tree for *Oreinotinus* (Fig. 1a, Methods, Supplementary Fig. 1 and Supplementary Notes 2 and 3), shows progressive movement from the north, in eastern Mexico and Oaxaca, to Chiapas and

Fig. 2 | Leaf form evolution in *Oreinotinus*. a, Upper left: NMDS morphospace showing the position of 39 *Oreinotinus* species scored for 3 quantitative and 2 categorical leaf variables (Methods, Extended Data Fig. 5 and Supplementary Fig. 4). Colours correspond to the four leaf ecomorphs: DEN, pink; LPT, green; SGE, blue; IGE, orange (see main text). The species present in each of the 11 areas of endemism are shown on the same NMDS plot, coloured by leaf ecomorph. E=East, N=North, S=South, W=West, Bol=Bolivia CA=Central America, Col=Colombia, CR=Costa Rica, Ecu=Ecuador, Jam=Jamaica, Mex=Mexico, Ven=Venezuela. **b**, Representative leaf silhouettes (Methods) for each species show the distribution of leaf forms on the RAxML species tree showing branch lengths and bootstrap values based on 100 replicates (black nodes >85%). Nine areas of endemism contain species with two or more ecomorphs, four areas contain three ecomorphs and one (Oaxaca) contains all four. **c**, Representative leaves: SGE, *V. acutifolium* (Oaxaca); DEN, *V. dentatum* (outgroup, eastern North America); IGE, *V. fuscum* (Oaxaca); LPT, *V. microcarpum* (eastern Mexico).

Central America, and from there once into the Caribbean and once into South America (~5 Ma). Species richness is higher in the longer-occupied areas and lower in more recently occupied regions, independent of area size (for example, seven species in Oaxaca versus two in Peru-Bolivia and Jamaica-Cuba).

Most importantly, species within most of the areas of endemism form clades (Figs. 1a and 2b and Extended Data Figs. 1 and 2),

that is, species in the same region are more closely related to one another than they are to species in other regions, consistent with in situ speciation. Oaxaca is an exception with two separate clades, while the Chiapas and Central America region is paraphyletic both with respect to Jamaica/Cuba and to the Costa Rica–South America clade (Figs. 1a and 2b). In striking agreement with these findings, dimensionality reduction methods (uniform manifold

approximation and projection for dimension reduction, UMAP; principal component analysis, PCA) applied to single nucleotide polymorphisms (SNPs) demonstrate the genetic similarity of species within each area of endemism and the genetic disparities among them (Fig. 1c, Methods, Extended Data Fig. 4 and Supplementary Note 4). Specifically, the genetic structure corresponds strongly to the 'north-to-south' layout of the geographic regions in our study, with three large distinct regions in Mexico, Central America and the Caribbean separated by a large gap from South America, within which several distinct clusters are also identified.

Leaf form evolution. As *Oreinotinus* diversified, several leaf forms evolved independently from an ancestral leaf form that seems to have facilitated repeated differentiation along particular axes of variation (parallel evolution, as used here; Fig. 2). We distinguish four leaf morphs in the field and herbarium. (1) A DEN form with medium-sized ovate-shaped leaves, with conspicuous, regularly spaced marginal teeth and sparse pubescence. The DEN form is shared with the *Dentata* clade of eastern North America[24,25] and is likely to be ancestral in *Oreinotinus* (Methods and Extended Data Fig. 6). (2) An LPT form with large leaves that are densely stellate pubescent on the abaxial surface and bear marginal teeth. (3) An SGE form with small and glabrous leaves with entire margins (that is, lacking teeth except for one or two pairs of glandular teeth (extrafloral nectaries[34]) near the intersection of the petiole). (4) An IGE form that is intermediate in size, with glabrous to sparsely pubescent lamina and entire margins.

These leaf forms differ in traits of functional importance. Leaf size and shape are well known to influence light capture, thermoregulation and the efficiency of gas exchange[35,36], and leaf size varies predictably along light, temperature and aridity gradients[37,38]. Like leaf size, the presence of trichomes on the leaf surface has multiple functional effects: pubescence influences leaf energy budgets, gas exchange, reflectance, wettability and protection from herbivores[39–42]. Species that bear teeth along their leaf margins are associated with mesic temperate forests[43] where teeth have been hypothesized to be a means of boosting photosynthesis early in the growing season[44], to be a biomechanical and hydraulic repercussion of evolving thin leaves[36] or to be an outcome of the extended development of leaf primordia inside of overwintering buds[45].

To assess the distinctness of our four leaf forms we carried out non-metric multidimensional scaling (NMDS) analyses using 39 species (including *V. dentatum* from eastern North America, but excluding *V. hirsutum* and *V. membranaceum*, for which we lacked adequate sampling; Methods) scored for five relevant leaf characters, including three quantitative variables (leaf blade area, leaf length relative to width (aspect ratio) and the density of marginal teeth) and two categorical variables (a four-state pubescence density character and a two-state trichome type character; Methods and Supplementary Table 5). These analyses identified one leaf form corresponding to our DEN category, one to our LPT category, and one including both our IGE and SGE categories (Extended Data Fig. 5 and Supplementary Fig. 4). Leaf sizes vary sufficiently in nature to produce some overlaps between the IGE and SGE forms, but we are able to consistently distinguish them in the field, in part based on additional subtle differences in leaf form and other characteristics not captured by the variables measured to date (for example, overall plant size: most SGE species are small, highly branched shrubs, whereas most LPT and IGE plants mature as small trees). For the species that we identify as IGE and SGE, we carried out an additional analysis of leaf blade area, which revealed strong statistical support for a bimodal distribution of leaf size (Extended Data Fig. 5). We also note that a PCA using the three quantitative traits alone, or also including a quantitative representation of the four pubescence states, yielded very similar results, with our four leaf forms largely occupying adjacent and non-overlapping
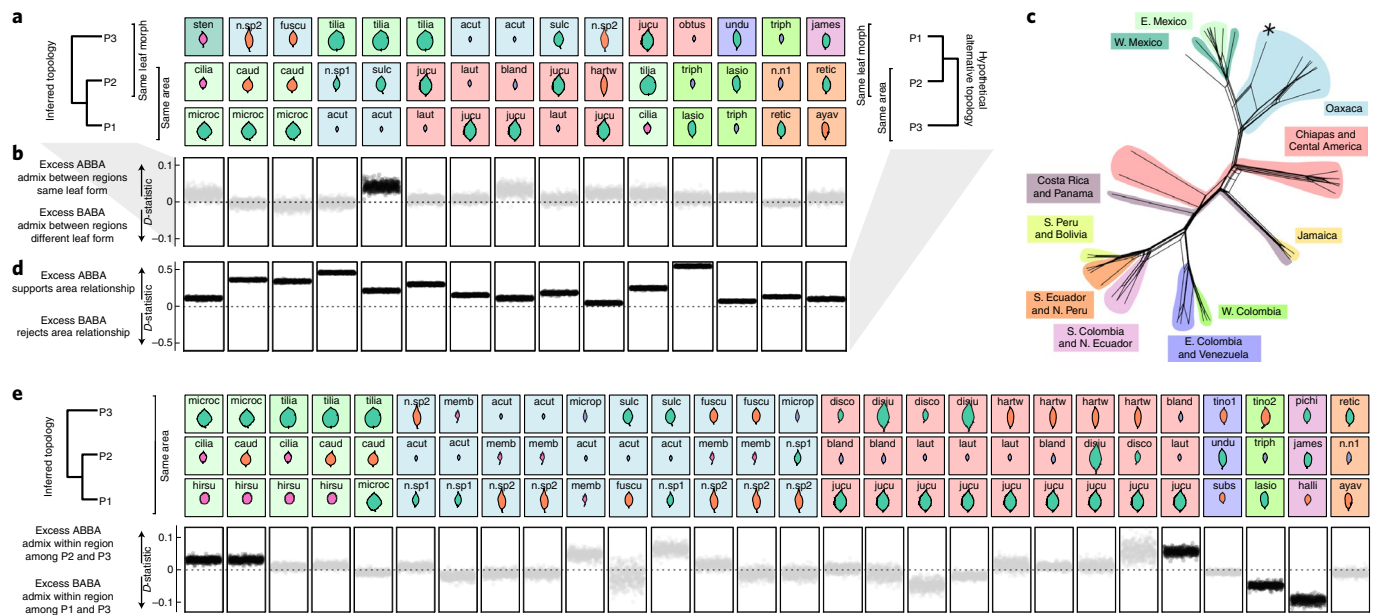
regions of PC morphospace (Methods and Supplementary Fig. 7). Furthermore, phylogenetic independent contrast analyses show strong correlations among the variables that define our leaf forms (Supplementary Fig. 6). On the basis of these several analyses we retained the distinction between IGE and SGE forms in further analyses, thus recognizing four leaf forms overall, as shown in the upper left-hand plot in Fig. 2a. The additional plots in Fig. 2a highlight the leaf forms of the species that occupy the 11 areas of endemism in the NMDS space. Representative leaf silhouettes for each species (Methods) are displayed on the species tree in Fig. 2b, and examples of the four leaf forms are shown in Fig. 2c.

Together, our phylogenetic, biogeographic and morphometric analyses clearly document a replicated radiation. The three derived leaf morphs did not originate early in *Oreinotinus* evolution and then spread among the areas. Instead, as lineages diversified within each area of endemism they repeatedly traversed the same regions of leaf morphospace. Of the 11 areas, 9 harbour at least 2 leaf morphs; 4 areas include 3 forms and 1 (Oaxaca) contains 4. Parallel evolution of the derived leaf forms is strongly supported by the exceptionally high levels of homoplasy in the leaf form character (Extended Data Fig. 6 and Supplementary Fig. 5). We also carried out analyses of the repeated occupation of particular regions in the phylomorphospace[46] and found that the derived leaf forms all show significant parallel evolution (IGE: $P = 0.005$; LPT: $P = 0.022$; and SGE: $P = 0.013$; Methods and Supplementary Fig. 7). Based on simulations using a posterior predictive framework (Methods), we reject a simple evolutionary model (in which leaf form and biogeographic range evolved independently) to explain this distribution of leaf forms among regions (Methods). Specifically, we find that it is unlikely by chance alone to observe, as we do, 9 areas of endemism with 2 or more leaf forms (0.7% or 1/150 null samples) and 5 areas with 3 or more forms (2% or 3/150 null samples; Supplementary Fig. 8). Furthermore, we devised a test to compare rates of trait evolution (Methods) and found that the Brownian rate of trait evolution is higher among species within a region than it is among comparably sized sets of randomly sampled species (Supplementary Fig. 9). Based on these several lines of evidence, we conclude that particular leaf forms have evolved in parallel within multiple regions and that leaf form diversity and rates of trait evolution within regions are greater than expected by chance.

Owing to the high level of homoplasy in leaf form across the phylogeny, the precise order in which different morphs evolved within regions is difficult to infer (Extended Data Fig. 6 and Supplementary Fig. 5). Parsimony inferred 22+ shifts in leaf form, with equivocal character states at the deepest nodes, but with strong support for the IGE form as ancestral to the South American radiation. The ancestral DEN ecomorph is not present south of the Isthmus of Tehuantepec; it appears to have been retained in the northern regions and does not exhibit significant parallel evolution[46] (DEN: $P = 0.064$). Very round leaves are also absent south of Oaxaca and there is an evident trend towards fewer marginal teeth, more elongate leaves and slightly smaller leaves in the South American radiation (Supplementary Fig. 5). The IGE form is the most common (16/40 species studied here; Supplementary Table 1) and perhaps the ancestral form south of the Isthmus of Tehuantepec and more certainly in the lineage that entered South America (Extended Data Fig. 6). We speculate that species with IGE leaves may have spread between regions, while the SGE and LPT forms evolved repeatedly in situ. This is consistent with our observation that IGE species generally have larger geographic ranges and inhabit lower elevations compared with LPT and SGE species, which generally have smaller ranges at higher elevations (see below and Extended Data Fig. 3).

Admixture, network analyses and reproductive isolation. Studies of adaptive radiations have increasingly revealed an important role for introgressive hybridization in the process (for example,
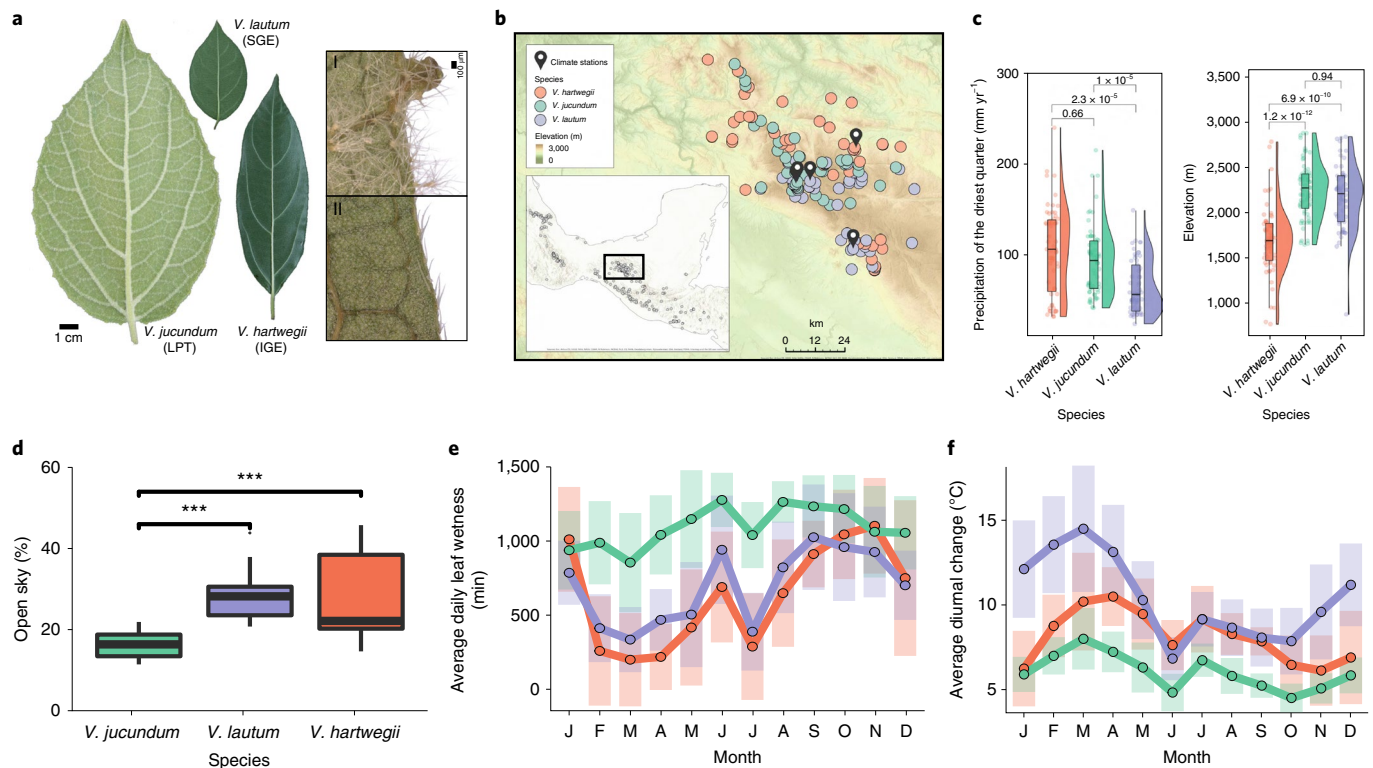
Fig. 3 | Admixture and network analyses. a, Selected ABBA-BABA tests using *V. dentatum* of eastern North America as an outgroup, and an ingroup with two species from the same area of endemism with contrasting leaf forms and one species from a different area with either form (Methods). Background colours correspond to areas in Fig. 1 and leaf silhouette colours to ecomorphs in Fig. 2; names in boxes correspond to species names in Figs. 1 and 2 (also see Supplementary Tables 1 and 2) **b**, Admixture is rare among species from different regions. Each test shows *D*-statistics for 500 bootstrap replicates, coloured dark if the test deviates from zero with a *Z*-score > 5. **c**, Phylogenetic network analysis shows correspondence with the species tree and areas in Figs. 1 and 2 (Supplementary Fig. 10). Only *V. sulcatum* (Oaxaca; asterisk) shows connections with species from another region (Extended Data Fig. 7). **d**, ABBA-BABA tests on a hypothetical alternative phylogeny in which species with similar leaf forms share a most recent common ancestor, showing significant discordance in all tests. **e**, Admixture among species within the same regions is uncommon.

in Darwin's finches[47], cichlid fishes[48] and *Heliconius* butterflies[49]). This, along with our field observations and analyses of localized hybrid swarms involving several *Orienotinus* species[22], prompted our exploration of the possible role of hybridization in this case. We used *D*-statistics (ABBA-BABA tests[50]; Methods) to examine SNP patterns among four-taxon subtrees, focusing on tests with species from neighbouring geographic regions with similar leaf forms (Fig. 3a). Tests performed on trees matching our inferred phylogeny show little evidence of admixture, except in *V. sulcatum* (Oaxaca), which appears to be admixed with the eastern Mexico clade (Fig. 3b and Extended Data Fig. 7). A phylogenetic network inferred from gene trees[51] (Methods) similarly identifies *V. sulcatum* as a rare case of inter-region admixture (Fig. 3c and Supplementary Fig. 10). We also investigated an alternative hypothesis in which species with similar leaf forms were assumed to share a recent common ancestor. Under this hypothesis, subtrees exhibit significant discordance (Fig. 3d), with an excess of allele sharing indicated among species from the same region. This implies either that this alternative hypothesis is false, or that there is significant gene flow among cohabiting species. *D*-statistics on subtrees of species from the same region clearly refute the latter, showing that admixture is generally rare within regions (Fig. 3e and Extended Data Fig. 7). We thus reject alternatives to replicated radiation: species with similar leaf forms in different regions are not admixed and do not appear to share recent ancestry, and species within regions show little sign of admixture that could potentially confound our phylogenetic analyses.

The flowers and fruits of all *Oreinotinus* species show only minor differences (for example, reduced endocarp grooving in South American species[52]), and we have succeeded in crossing individuals of several Mexican species with different leaf morphs[19] (all of them known tetraploids of $2n = 36$; ref. [22]). Although hybridization is possible, differences in flowering times of co-occurring species (with adjacent or overlapping geographic ranges) appear to maintain

reproductive isolation in the field[22], as they do in *Viburnum* elsewhere[53,54]. Flowering times sometimes, however, overlap in contact zones where habitats and light regimes have been altered by human activity, and localized hybrid swarms can occur, as documented for *V. lautum* (SGE) and *V. jucundum* (LPT) in Chiapas[22]. Yet, as shown above, we find little evidence of introgression outside these hybrid swarms and, with the exception of *V. sulcatum* (see above), no evidence of hybridization involving species from different areas (Fig. 3 and Supplementary Fig. 8).

Niche differentiation. To explore the adaptive value of different leaf forms, we characterized the environments of *V. hartwegii* (IGE), *V. jucundum* (LPT) and *V. lautum* (SGE) on the Central Plateau of Chiapas, Mexico, where the ranges of these species overlap (Fig. 4a). We monitored climate and light variables over 3 years using portable weather stations installed at multiple sites for each species (Methods and Fig. 4b) and measured canopy openness (Methods and Fig. 4d). These data reveal that *V. lautum*, with its small glabrous leaves (SGE), experiences higher irradiance, greater daily temperature fluctuations, occasional freezing temperatures and less cloud cover during the drier winter months in comparison with *V. jucundum* with its large pubescent leaves (LPT; Fig. 4e,f). *V. hartwegii*, with glabrous leaves of intermediate size (IGE), generally grows at lower elevations on the flanks of the Central Plateau, at sites experiencing more precipitation and less diurnal temperature change during the winter than *V. lautum* (Fig. 4c,e,f). Climatic data extracted from georeferenced occurrences on the Chiapas plateau are consistent with these findings (Fig. 4b,c; and Extended Data Fig. 8), showing that *V. lautum* occupies significantly drier sites than *V. jucundum* and that *V. hartwegii* grows at lower elevations with higher temperatures throughout the year. These results are well supported despite the fact that the ranges of *V. lautum* and *V. jucundum* broadly overlap surrounding the valley of San Cristóbal de las Casas

Fig. 4 | Ecological differentiation of species with different leaf forms in Chiapas, Mexico. a, Representative leaves of *V. jucundum* (LPT), *V. lautum* (SGE) and *V. hartwegii* (IGE). Leaf margin images of *V. jucundum* (I: dense pubescence of stalked stellate trichomes on the blade and margin, showing one tooth) and *V. lautum* (II: glabrous, with simple trichomes on margins only, lacking teeth). **b**, Collection localities for niche analyses and the location of climate stations on the Central Plateau of Chiapas, Mexico. **c**, Environmental data from georeferenced collection localities (Methods) show *V. lautum* (*N* = 77) in drier sites and *V. hartwegii* (*N* = 73) at lower elevations than *V. jucundum* (*N* = 82). All CHELSA climate variables for these three Chiapas species are displayed in Extended Data Fig. 8. Box plot elements: centre line, median; box limits, upper and lower quartiles; whiskers span 1.5× the interquartile range. Density ridgeline plots drawn using the geom_density_ridges function in the ggridges R package[94]. Vertical line within each density ridge represents the mean value for that region. *N* = 181 biologically independent samples (59 *V. hartwegii* specimens, 65 *V. jucundum* specimens, 57 *V. lautum* specimens). **d**, Canopy openness estimated using fisheye lens photography (Methods) shows that *V. jucundum* (*N* = 20) occupies more closed forests than *V. lautum* (*N* = 20) and *V. hartwegii* (*N* = 10); *** p<0.001, box plot elements are as in **c**. **e**, Climate station data show *V. jucundum* exposed to more consistent fog, especially during the dry season (February–May); coloured bars for each month: one standard deviation above and below the mean (dot). **f**, Climate station data show *V. lautum* experiencing higher daily and seasonal temperature fluctuations; temperatures at *V. jucundum* sites are more stable throughout the day and year; coloured bars as in **e**. Credit: leaf margin images in **a**, J. Negvesky, Keyence Corp.

(Fig. 4b), probably due to long-term human disturbances that have brought populations of these two species into contact and promoted localized hybridization[22].

Although we have not been able to compare the niches of ecomorphs in the other areas of endemism in the same detail as in Chiapas, our analyses of georeferenced collection localities from throughout the range of *Oreinotinus* (Extended Data Fig. 3) provide support for the hypothesis that the leaf ecomorphs sort into climatic niches similar to the ones we have characterized in Chiapas. Despite the confounding effects of major latitudinal and elevational differences among the 11 areas of endemism (Supplementary Fig. 11), IGE species across the entire range are similar to *V. hartwegii* in occupying generally lower elevations and experiencing higher temperatures and precipitation throughout the year, with less extreme diurnal temperature ranges and seasonality (Extended Data Fig. 9 and Supplementary Fig. 12). Likewise, as in our comparison of *V. jucundum* and *V. lautum*, SGE species differ from LPT species across the range of *Oreinotinus* in experiencing less precipitation throughout the year (both in the warmer wetter season and the cooler drier season) and greater seasonal and diurnal temperature ranges (Extended Data Fig. 10). Comparing ecomorph niches region by region, the results are again broadly consistent with our

findings for the three Chiapas species; where we recover significant differences among different leaf forms in regions where all three forms occur, these generally (but not always) conform to our expectations (Supplementary Figs. 13, 14 and 15).

The differences in climate niches suggest several adaptive hypotheses for the origination of the SGE and LPT ecomorphs. The small, glabrous leaves of *V. lautum* and other SGE species would result in a thinner leaf boundary layer that would allow for a closer tracking of leaf and air temperatures. This would prevent overheating on warm sunny days and would also prevent overchilling when temperatures occasionally dip below freezing[38]. The large, pubescent leaves of *V. jucundum*, and presumably other LPT species, correlate with a lower light, frequently cloudy environment, with water condensing on leaf surfaces on most days. We hypothesize that the larger leaf size aids in light capture[37], while the dense abaxial trichome covering aids in gas exchange by repelling water from the leaf surface[41]. We are uncertain about the functional role of leaf teeth in the *Oreinotinus* radiation, but we note the lack of the DEN ecomorph south of Oaxaca, Mexico, and a decline in tooth density, even in the LPT ecomorph, south of the Isthmus of Tehuantepec and especially in South America (Extended Data Fig. 6 and Supplementary Fig. 5). We suspect that the distribution of marginal teeth may largely reflect

the ancestry of the *Oreinotinus* lineage in temperate forests of North America. The reduction in teeth in the LPT ecomorph and the loss of teeth in the SGE and IGE forms may reflect shifts in phenology and a reduction in leaf development within resting buds with adaptation to less seasonal environments[45]. Although more detailed comparisons are necessary across the entire range of *Oreinotinus*, our results indicate that the leaf ecomorphs are adapted to specific sets of subtly different, often adjacent, environmental conditions. Leaves are, of course, parts of integrated whole plants, and differences in other possibly correlated and relevant traits (for example, plant size and wood anatomy) could also contribute to the fitness of ecomorphs in different environments.

Conclusions

In summary, we have documented a case of the replicated evolution of a set of leaf ecomorphs across disjunct mountainous regions in the neotropics. Our example has played out at a continental scale, but in circumstances that mirror an island archipelago. And, as in some island systems, our cloud forest regions were occupied at different times as *Oreinotinus* spread southwards, providing the opportunity to study diversification at different temporal stages.

The rapid diversification of *Orienotinus* displays elements of non-adaptive and adaptive radiations[55]. We envision that in the early non-adaptive stages, as the lineage spread southwards, allopatric separation yielded species that were largely ecological replacements of one another. However, later in the process, as speciation occurred within the disjunct regions, there were repeated adaptive shifts in leaf form.

Many questions remain concerning the developmental underpinnings of the leaf traits and exactly how these function in relation to relevant environmental variables. However, the analyses reported here firmly establish a case of replicated radiation in plants that rivals those in animals, which we predict will likewise provide deep insights into speciation and functional diversification.

Methods

RAD-seq data. Leaf tissues were collected from herbarium samples or newly collected plants (Supplementary Tables 2 and 3), including multiple specimens from different geographic locations whenever possible. DNA was extracted using Qiagen DNeasy plant extraction kits. RAD-seq data were generated by Floragenex by digesting genomic DNA with the PstI restriction enzyme followed by sonication and size selection at 400 base pairs (bp). Samples were ligated with 10 bp barcodes for multiplexing. The 180 samples used in this study (Supplementary Table 3) were pooled with other *Viburnum* samples and sequenced on Illumina HiSeq 2500 or 4000 to produce 100 bp single-end reads. Genomic data were demultiplexed to samples and assembled using ipyrad v.0.9.65. All further analyses and assembly statistics pertain only to the 180 samples used in this study. These samples recovered 3.73 million ± 3.26 million (mean ± standard deviation (s.d.)) reads per sample. Genomic reads were assembled into orthologous loci using the reference-based assembly method in ipyrad. Default settings were used to trim and filter reads prior to mapping to the draft reference genome of *V. lautum* (see below). Reads mapping to multiple genomic locations were discarded as paralogues. The proportion of mapped reads varied widely across samples (mean = 76.9%, s.d. = 18.4%), but did not correspond strongly with species identities; variation in the quality of DNA extractions probably explains the variation. Coverage was relatively even among clusters of mapped reads, such that the total number of clusters within samples (mean = 76,317, s.d. = 23,849) was similar to the number that passed our minimum depth threshold of 6 (mean = 68,530, s.d. = 21,876). When calling consensus alleles we allowed for up to 4 alleles per sample; all species of *Oreinotinus* counted to date are tetraploids of $2n = 36$, but ploidy is not known for all *Oreinotinus* species[27]. The final assembled dataset comprised 201,009 loci, representing non-overlapping genomic regions where data from at least 4 samples mapped to the reference genome and passed all filtering. Each sample on average is present in 60,246 loci (s.d. = 18,584). This dataset represents the largest but most sparse assembly of the RAD-seq data. Rather than compile multiple assemblies of the dataset using different parameter settings to allow for different proportions of missing data, we instead created only this largest assembly and performed downstream analyses using the ipyrad-analysis toolkit[56], which allows filtering and subsampling loci, genomic windows, or SNPs into multiple file formats for specific downstream analysis tools. These filtering options are described in the phylogenetic inference, PCA and ABBA-BABA sections below.

Reference genome assembly. Flash frozen young leaf tissues were collected from an individual of *V. lautum* cultivated from a sample of a wild population near the town of Teopisca, Chiapas, Mexico (YAL1123). A voucher specimen of this population is deposited in the Yale University Herbarium (PWS 3105). Tissues were sent to Dovetail Genomics for high-molecular weight DNA extraction and sequencing. Five Illumina libraries were prepared and sequenced to generate paired-end 150 bp reads. Adapter and quality trimming was performed using trimmomatic[57], with the ILLUMINACLIP mode used to remove sequencing adapters, and bases with quality scores <20 removed from the leading and trailing ends of reads. Paired reads had variable insert sizes ranging from 322 to 568 bp and summed to a total of 3.15 billion read pairs, representing 917 Gb of sequence data after adapter and quality trimming.

A de novo draft genome was assembled from short read data using the Meraculous2 software[58] in diploid mode with *k*-mer size set to 73 and minimum *k*-mer frequency to 12. Meraculous is a whole genome assembler for Eukaryotic genomes that uses a hybrid *k*-mer/read-based method to assemble the genome into non-redundant haplotigs, or diplotigs when variation is present. The order and orientation of contigs, and the closing of gaps between them, is performed by mapping paired-end reads to the initially constructed contigs. The genome size of *V. lautum* was estimated from *k*-mer frequencies in the shotgun sequence data using GenomeScope 2.0[59] as 6.293 Gb. The de novo genome assembly yielded 542,881 contigs representing 3.0395 Gb in length. This represents an estimated 48% of the total genome size and 76% of the non-repetitive genome size. Scaffolding from short-read data reduced this to 482,845 scaffolds, with an N50 of 11.5 Kb and L50 of 70,032 scaffolds.

To further scaffold the de novo assembly, three long-range mate pair Chicago libraries were prepared by Dovetail Genomics and sequenced to generate 2x150bp reads. The three Chicago libraries had insert sizes ranging up to 300 Kb in length, and totalled an additional 1.25 billion read pairs. Scaffolding analysis was performed by Dovetail Genomics using the de novo shotgun assembly and Chicago data as input to HiRise (version three-pass-8cd24cd4b80d), a software pipeline designed specifically for using proximity ligation data to scaffold genome assemblies[60]. Finally, two Hi-C libraries were prepared for long-range scaffolding, producing an additional 212 million paired-end reads spanning 10–10 kb insert sizes and estimated to provide 23,864.92× physical coverage of the genome. Dovetail Hi-C library sequences were aligned to the Chicago scaffolded input assembly using a modified SNAP read mapper (http://snap.cs.berkeley.edu). The separations of Dovetail Hi-C read pairs mapped within draft scaffolds were analysed by HiRise to produce a likelihood model for genomic distance between read pairs, and the model was used to identify and break putative mis-joins, to score prospective joins and make joins above a threshold. Scaffolding with Hi-C made 79,674 joins, improving the scaffold N50 to 7.3 Mb and N90 to 92 Mb. Half of the genome is present in the largest 121 scaffolds (L50). The annotation of this draft genome and comparisons to other published genomes will be published elsewhere.

Individual-level tree inference. We reconstructed a maximum likelihood (ML) phylogeny from concatenated sequences to produce an 'individual specimen level' tree that includes 180 samples from our complete dataset (Supplementary Table 3). These samples were selected to exclude any putative hybrids. Based on our past and ongoing studies of highly localized hybrid swarms involving several *Viburnum* species in Puebla, Oaxaca and Chiapas[22], we are able to accurately identify possible hybrid plants by unusual combinations of leaf, twig and inflorescence traits. All specimens from hybrid swarms were omitted, as were several plants collected at some distance from these populations. The concatenated sequence alignment was produced using the ipa.window_extracter module in the ipyrad-analysis toolkit[56]. For computational speed we extracted data only from the 10 largest scaffolds and filtered sites to only maintain sequences shared by at least 25% of all samples and to exclude any samples that had less than 10% data present. The filtered matrix is 1.8 Mb in length, containing 109,825 SNPs and 50.7% missing cells. This alignment was used in RAxML[61] under the GTRCAT evolutionary model to infer a best tree under the rapid hill-climbing algorithm with support values from 100 bootstrap replicates. The tree was rooted by reference to *V. dentatum* samples from eastern North America[24,25]. Extended Data Fig. 1 shows that accessions from multiple populations of the same morphology-based species form generally well-supported clades in all but the five instances discussed in Supplementary Note 1.

Species-level *Oreinotinus* phylogeny. To build a species-level phylogeny, we selected one representative specimen from each of 41 species-level lineages (Supplementary Tables 1 and 2), and then extracted the 100 largest scaffolds using the module ipa.window_extracter in ipyrad-analysis[56]. Before concatenating all scaffolds, these were filtered, selecting only sites shared by at least 25% of all species. The resulting matrix is 7.6 Mb with 161,791 SNPs and 48% missing data. We ran RAxML[61] on this matrix using GTRCAT to infer a best tree under the rapid hill-climbing algorithm with support values from 100 bootstrap replicates. We used Toytree[62] to draw the resulting tree (Fig. 2b). This concatenation topology very closely matches the tree topology inferred using the multispecies coalescent methods described next.

Using ipa.treeslider included in ipa[56] we traversed the largest 100 scaffolds in our data in 643 non-overlapping windows of 2 Mb. We used a consensus reduction

approach to merge all specimens of each species to end with one representative for each of the 41 species (including 2 accessions for *V. tinoides*; see above). We reconstructed trees for each window using RAxML[61]. For the multispecies coalescent analysis, we selected only the 356 trees that contained more than 300 SNPs. Those trees were concatenated and used as input to ASTRAL-III v.5.7.1[63] (Extended Data Fig. 2).

Standard Bayesian divergence time methods encounter statistical and computational challenges when applied to RAD-seq datasets. In particular, the large numbers of short anonymous RAD loci are not easily combined into phylogenetic partitioning schemes. To circumvent these issues, we time-calibrated the species-level concatenation tree using RelTime[64] in MEGA X[65] (Fig. 1a). RelTime reapportions molecular branch lengths into branch-specific components of geological time and evolutionary rates, subject to user-provided constraints. For our RelTime analysis, we rooted *Oreinotinus* on *V. dentatum* (eastern North America) and limited molecular rate variation to <20-fold differences across branches. We additionally restricted the possible crown age of *Oreinotinus* to lie between 6.66 Ma and 14.82 Ma, based on the 95% highest posterior density (HPD95) for the *Oreinotinus* crown age obtained previously[22]. Our RelTime node age estimates for *Oreinotinus* were highly congruent with previous node age estimates obtained for the clade using a fossil-based Bayesian dating strategy[24]. This congruence was assessed using tests described in Supplementary Fig. 1 and Supplementary Note 2.

Oreinotinus map and areas of endemism. We combined our own collection localities with all *Oreinotinus* localities attached to preserved specimens available in the Global Biodiversity Information Facility (GBIF) (Extended Data Fig. 3). We filtered these specimen records by searching for and downloading localities for recognized species, along with known synonyms (for example, 'rhombifolium' in addition to 'tilaefolium') and potential misspellings (for example, 'tilliafolium'). We filtered the GBIF data further using the R package CoordinateCleaner[66] to remove points with a high probability of error (for example, localities with equal latitude and longitude, points in the ocean, localities with coordinate uncertainties greater than 1 km). We also removed duplicate points where the same species was recorded multiple times at an identical latitude and longitude. Finally, we consulted expert range maps and descriptions for *Oreinotinus* species to ensure that localities retained for each species were consistent with any known species ranges. In cases of uncertain identifications, we examined records with images in GBIF to properly assign specimens under our current taxonomy (Supplementary Tables 1 and 2).

We inputted the full dataset of 3,034 specimen records into ArcGIS Pro, plotted the points and uploaded the final map of species ranges to the Yale University ArcGIS online repository (https://arcg.is/1OLDXD0). Although our sampling is thorough with regard to available databases and collections, it is unlikely to cover the full ranges of *Oreinotinus* species. Based on our biogeographic analysis and the existence of contiguous mountain ranges in some areas, apparent disjunctions within particular areas of endemism (as in northern and central Peru) are probably an artefact of under-collection, rather than the true absence of *Viburnum* in suitable habitats in such gaps. However, the absence of *Viburnum* collections in low-lying areas between disjunct mountainous regions, such as in the Isthmus of Tehuantepec or the Isthmus of Panama, represent genuine gaps in species ranges and true barriers to dispersal between the cloud forest habitats occupied by these plants.

To delimit geographical areas of endemism that were based initially on topographic maps and our own field studies, we analysed our locality data for the full range of *Oreinotinus* using the online software Infomap Bioregions[67] (Supplementary Fig. 2). Infomap Bioregions implements a bipartite network and clustering algorithm to elucidate underlying structure in the data, partitioning geographic space into bioregions based on the distributions and species richness of the focal taxa. When preparing the dataset for this analysis, we retained only one randomly selected point for each species within each 1 km² grid cell to minimize the effects of sampling bias and spatial autocorrelation. This dataset and the ASTRAL species tree were loaded into the software. Specimen occurrence records were binned into cells using an adaptive resolution method such that each cell was no more than 4 × 4° and no less than 1 × 1°, and each cell contained no more than 100 records and no less than 10. These parameter values (the defaults provided by the software) are deemed most likely to reduce both overfitting and underfitting, given the scale of our dataset (3,034 species occurrence records).

This clustering approach identified ten bioregions, each with endemic species, and in most cases forming clades in our ASTRAL tree (Supplementary Fig. 2). Only four species occur in more than one of these areas: *V. microphyllum*, *V. obtusatum*, *V. tinoides* and *V. toronis* (Supplementary Table 1 and Supplementary Fig. 3). Of these, we note that *V. tinoides* could be considered to be two species, in two separate regions, based on the placement of accessions in two separate clades (Figs. 1a and 2b, Extended Data Figs. 1 and 2 and Supplementary Note 1): *V. tinoides* (1) accessions are linked with species from western Colombia and *V. tinoides* (2) accessions are linked with species from eastern Colombia. However, because we are unable to separate accessions from these two clades based on morphological traits or ecological variables, we have retained *V. tinoides* in the broad sense. The resolution of this taxonomic problem requires more extensive sampling throughout the range, along with additional morphological and ecological analyses.

The only region that we consider to be an area of endemism that was not recovered in our Infomap Bioregions analysis is Oaxaca (labelled area 3 in Supplementary Fig. 2), which clustered with eastern Mexico. We recognize Oaxaca as one of our 11 areas of endemism based on the fact that 6 *Viburnum* species are endemic to this small region (*V. acutifolium*, *V. fuscum*, *V. membranaceum*, new species 1, new species 2 and *V. sulcatum*), more than in any of the other areas. This region is also recognized as an area of endemism for other montane plant groups and is separated from eastern and central Mexico in analyses of geology and vegetation (for example, ref. [31]). Specifically, we note a division between these areas south of Huatla, Oaxaca, and marked roughly by the Río Santo Domingo drainage (near San Bartolomé Ayautla). The failure of Infomap Bioregions to identify Oaxaca as a separate area probably stems from the under-sampling of *Viburnum* in that region (which has the fewest occurrence records per species of any of the areas of endemism) and the exceptionally narrow known ranges of each of the endemic species we recognize in that area.

Oreinotinus historical biogeography. We reconstructed ancestral species ranges for *Oreinotinus* using the DEC approach[33] (Fig. 1a). DEC models how species evolved among a set of discrete regions along the branches of a phylogeny through an inferred sequence of dispersal, (local) extinction and cladogenetic events. DEC models require as input a time-calibrated species-level phylogeny and discrete range data (that is, inhabited sets of discrete regions) for the species of interest. For the phylogeny, we used the dated RelTime tree described above (Fig. 1a). Discrete range data were assembled using new geo-coordinates from this project, curated GBIF coordinates and expert range maps (see above).

We designed a phylogenetic model in RevBayes[68] that reconstructed the historical biogeography of *Oreinotinus* under the DEC model[33,69]. We assumed the topology and divergence times of the phylogeny were identical to those of the ML tree. We assumed all dispersal rates were equal, with a prior distribution of Exponential(10), and all extirpation rates were equal, also with a prior of Exponential(10). Within-region ('sympatric') cladogenetic events occurred with prior probability $u \sim$ Uniform(0,1) and between-region ('allopatric') cladogenetic events occurred with the probability $1 - u$. We ran the Markov chain Monte Carlo (MCMC) analysis for 3,000 RevBayes iterations, with 10 moves per iteration. All parameters had effective sample sizes of >200. Ancestral ranges were sampled during MCMC then summarized and plotted back against the species-level time-calibrated phylogeny reference tree. Phylogenetic uncertainty did not impact the major features of our primary ancestral range estimates (Supplementary Fig. 3).

PCs and UMAP analyses of SNP data. SNPs were extracted from the HDF5 database file produced by ipyrad using the ipa.pca module in ipyrad-analysis[56] and filtered separately in each of several different analyses to minimize missing data among the set of subsampled taxa included in each run. The analyses presented in the main text (Fig. 1c) used filter settings to retain only SNPs that have a minimum coverage of 75% across the subset of samples (mincov=0.75), and a minimum coverage of 50% among the samples within each circumscribed region in the analysis (minmap=0.5), or of 1 in the case of the full set of samples (minmap=1), and required a minor allele frequency greater than 2% (minmaf=0.02). As missing data are a common feature of RAD-seq data but are not accommodated in dimensionality reduction methods such as PCA, we imputed missing values after filtering using the 'sample' imputation method in ipa.pca. This randomly samples two alleles at each missing genotype based on the frequency of derived alleles at the missing site among other individuals in the population (group) to which a sample is assigned. In our main analysis, every sample was assigned to 1 of 11 populations corresponding to the geographic areas of endemism defined in this study. To reduce potential biases caused by a priori assignment of samples to population groups used in imputation, we also performed analyses using a *k*-means clustering approach to assign samples to groups. This method first performs a PCA on a highly filtered set of SNPs allowing for little missing data (for example, mincov=0.90), then uses *k*-means clustering to assign samples to *k* groups based on their distances in multidimensional PC space. The threshold for filtering SNPs is then lowered, allowing for additional SNPs in the dataset, for which missing values are then imputed again using the 'sample' method but now using the *k*-means-based population assignments. This is repeated iteratively allowing for more and more data to be included, and imputed, until the target mincov setting (here mincov=0.75) is reached.

With the filtered and imputed SNPs extracted by the ipa.pca module, we ran either a UMAP[70] or PCA[71] on the data, which ipa.pca implements from scikit-learn[72]. The former is a nonlinear dimensionality reduction method that often performs better than PCA for visualizing highly dimensional data in two dimensions, by capturing both local and global structure. Both UMAP and PCA were performed on the full dataset, as well as on subsets of the data that focused only on selected samples from smaller regions (for example, Mexico, Central America and the Caribbean or South America). To reduce the effects of linkage, only a single SNP is subsampled per RAD locus. For the PCA and UMAP plots shown in Fig. 1, the 'All Regions' analysis includes 178/180 samples classified in the 11 previously defined regions (outgroups excluded). The post-filtered dataset has 5,497 SNPs from 1,843 distinct RAD loci with 21% of data imputed. The 'Mexico and Central America' analysis includes 116 samples with 2,931 post-filtered SNPs

from 823 linkage blocks and 21% missing data. The 'South America' analysis includes 62 samples with 37,737 post-filtered SNPs from 11,006 linkage blocks and 19% missing data. Additional PCA and UMAP analyses are presented in Extended Data Fig. 4 (also see Supplementary Note 4); these demonstrate the limited effects of subsampling SNPs, imputing missing data and assigning populations a priori.

Leaf morphological analyses. Digitized leaf specimens were assembled for all *Oreinotinus* species from various sources, including herbarium sheets and standardized images taken by us in the field. Although we obtained morphometric data for *V. hondurense* (Honduras), *V. venustum* (Costa Rica) and *V. incarum* (Peru), these were excluded from our statistical analyses because we were unable to obtain DNA of sufficient quality to include them in our RAD-seq analyses. *V. hirsutum* and *V. membranaceum* were included in our phylogenetic analyses but were excluded from our morphometric studies because we were able to obtain leaf measurements for fewer than five plants that poorly represent variation across their geographic ranges. Specimens showing any evidence of hybridization between species were omitted (see above). In view of documented seasonal heteroblasty in *Viburnum*[73], leaf measurements were taken exclusively from leaves situated immediately below inflorescences whenever possible. One leaf per specimen was selected for measurement based on an unobstructed view of its profile. For species represented by few specimens, we used several leaves from a single individual. The final dataset consisted of 657 individual leaves representing 647 individual plants, with an average of $N = 15$ leaves per taxon (range = 7–29) (Supplementary Table 5).

Digital measurements were made in ImageJ[74] with the assistance of the plugin LeafJ[75]. To calibrate measurements correctly, for each image the pixel/cm ratio was established using either a ruler in the image frame or another known measurement from the image. Owing to LeafJ's input requirement of a leaf without background, selected leaves were lassoed from the specimen images (with pixel ratio maintained) with the assistance of a photo editing program. LeafJ also requires a petiole to orientate automatic leaf measurements; in the few cases where the selected leaf image did not have a visible petiole, one was inserted at the base of the leaf for use in LeafJ, and LeafJ's measure of petiole length was discarded. LeafJ provided automated measurements of blade length, width, area and perimeter. Aspect ratio was calculated from blade length/blade width. The total number of marginal teeth was counted and divided by the total blade perimeter to obtain tooth density. Using the Python library scikit-learn[72] and its module preprocessing.StandardScaler, we standardized the variables to reduce bias by scale[76], transforming all data to have mean = 0 and variance = 1.

As trichome density could not be accurately measured from our digital images, we used two categorical variables. One character, with four states, reflects the density of pubescence on the abaxial leaf surface: (1) glabrous or with a few trichomes on the major veins; (2) sparsely pubescent on the leaf lamina; (3) densely pubescent but with the leaf lamina visible beneath the trichomes and hence the surface generally appearing grey or brown in colour; and (4) densely pubescent with several matted layers of trichomes that completely obscure the lamina beneath, and hence appearing white in colour. A second, two-state character reflects a difference in the type of trichomes produced: (1) simple or bifid or rarely small unstalked stellate trichomes with few rays and (2) multicellular stalked stellate trichomes with multiple rays. Even mostly glabrous forms could be scored for this character as they produce a few trichomes either on the leaf margins, or at the base of major veins or on the petiole or in domatia in the axils of the major secondary veins.

Because our ecomorphs are characterized by a mix of continuous and categorical variables, we were unable to use the SURFACE approach to identifying leaf forms and convergence[77]. Instead, we conducted NMDS analyses using the modules manifold.MDS and cluster.KMeans in the library scikit-learn[72]. We used the elbow method to identify the optimal number of components to represent our data in the NMDS; this entails multiple iterations of the test varying the number of components and determining the point (elbow of the curve) where the test does not return a significantly better result (Extended Data Fig. 5). In this case, we used the metric stress (difference between distances in the reduced dimensions) to find the optimal number of components. The same elbow method was used to obtain the optimal number of *k* in the clustering by using the metric inertia (sum of distances of samples to their closest cluster centre) to measure the optimization. Both analyses, dimensional reduction and clustering, were done using a matrix with the means by species (Fig. 2a and Extended Data Fig. 5) as well as a matrix with information for more than 600 individual accessions (Supplementary Table 5 and Supplementary Fig. 4).

Clustering with $k = 3$ identified one cluster that included both our SGE and IGE ecomorphs (Extended Data Fig. 5). We further explored the distribution of leaf area in this cluster by comparing the means for the IGE and SGE forms with a Wilcoxon test implemented in the package ggpubr with the function stat_compare_means[78] in R. As shown in Extended Data Fig. 5c, this recovered a significant difference between the IGE and SGE forms in leaf area.

We also carried out PCA using our three continuous variables: blade area, aspect ratio and tooth density. Using the Python library scikit-learn[72] and its module preprocessing.StandardScaler, we standardized the variables to reduce bias by scale[76], transforming all data to have mean = 0 and variance = 1. We computed all principal components axes using the module decomposition.PCA

in the aforementioned library. To reflect trichome information in this analysis we used a digital microscope to manually count trichomes in 0.3 or 0.5 cm circular apertures on a standard sector of the adaxial surface and calculate trichomes cm^{-2} 150 specimens from the Yale herbarium selected from 34 species spanning the full spectrum of pubescence. From this, we obtained average trichome densities for our four pubescence density states as follows: 1 = 4.97, 2 = 107.07, 3 = 973.72 and 4 = 4466.34 trichomes cm^{-2}. The result of this analysis is similar to those obtained using NMDS, as shown in Supplementary Fig. 7a. Species PC scores from this analysis were used in analyses of parallel evolution (see below).

Using the time-calibrated species-level *Oreinotinus* phylogeny we performed ancestral state reconstructions of the four-state leaf form character using both parsimony and stochastic mapping approaches (Extended Data Fig. 6). We simulated 1,000 stochastic mappings based on an 'all rates different' character transition matrix estimated by ML, with equal root node prior probabilities. Results of the simulations were summarized as pie charts (Extended Data Fig. 6). Similar to parsimony and due to the high level of homoplasy, ancestral state estimation along the backbone was equivocal, in this case including the ancestor of the South American radiation. We also performed ML reconstructions of our three continuous variables (Supplementary Fig. 5) and calculated phylogenetic independent contrasts for logged values of aspect ratio, tooth density and blade area and implemented linear models of contrasts between all three combinations of the characters (Supplementary Fig. 6). All three characters exhibited correlated evolutionary change across the tree: leaf area and tooth density are positively correlated (Supplementary Fig. 6a), and tooth density and aspect ratio (Supplementary Fig. 6b) as well as blade area and aspect ratio (Supplementary Fig. 6c) are negatively correlated. All analyses were performed in R, using phytools[79] and castor[80] packages.

To further explore the extent of parallelism/convergence we used the convevol package for R[46,81] to assess the significance of the convergent evolution of the four leaf ecomorphs. We executed the function convnumsig, which simulates evolution under a Brownian motion model along our species-level tree, using parameters from the first two principal components from our PCA analysis including the semi-quantitative pubescence character (see above; Supplementary Fig. 7). For each leaf type, the simulated values were compared with the convnum metric to determine the significance of the convergence. This metric refers to the number of times a lineage has entered a region in a given morphospace. For this analysis, we ran 1,000 simulations and used the two most informative principal components. Supplementary Fig. 7 displays these results in the phylomorphospace.

The tree in Fig. 2b shows a representative leaf silhouette for each species. This specimen was defined as the closest in Euclidean distance to the species' centroid in principle components space.

Posterior prediction analysis of leaf type and biogeographical evolution. Of the 11 areas of endemism, 9 contain species of 2 or more leaf types ($N_2 = 9$), and 5 regions contain 3 or more leaf types ($N_3 = 5$). We tested whether the observed values for N_2 and N_3 in *Oreinotinus* exceeded null expectations using a Bayesian posterior prediction[82]. We defined a simple null model that extends our phylogenetic model of historical biogeography to also model leaf type evolution under an equal-rate four-state continuous-time Markov process with an Exponential(10) rate prior. By design, this null model neither allows leaf type to influence species range evolution nor species range to influence leaf type evolution. After inferring the joint posterior distribution of all biogeographical and leaf type evolution parameters under this null model, that same inference model simulated one dataset per sample of posterior parameter estimates. We removed eastern North America and *V. dentatum* from each simulated dataset to allow a valid comparison to our observed values of N_2 and N_3 in *Oreinotinus*. Finally, we computed values of N_2 and N_3 for all simulated datasets to generate a null distribution of regional leaf type variation to compare with those values for *Oreinotinus*.

Test of regional rates of trait evolution. We tested whether leaf traits evolved faster among geographically sympatric *Oreinotinus* lineages. We extracted 11 subphylogenies with $(n_1, n_2, \ldots n_{11})$ taxa from the *Oreinotinus* RelTime phylogeny for each of the 11 regions and then estimated the Brownian rate of trait evolution jointly across all subphylogenies for each trait (log blade area, log tooth density, aspect ratio) using ML. We then similarly estimated Brownian rates of trait evolution across 100 replicate datasets, each with 11 randomly sampled subphylogenies having $(n_1, n_2, \ldots n_{11})$ taxa. The finding of higher rates using regional subphylogenies relative to random subphylogenies supports a scenario in which geographic sympatry accelerates trait evolution (Supplementary Fig. 9).

D-statistics/ABBA-BABA tests. We used our genome-wide sampling of SNPs to test for admixture among species from neighbouring regions and for gene flow among species within regions (Fig. 3, Extended Data Fig. 7 and Supplementary Table 4). This was done using a series of D-statistic (ABBA-BABA) tests[50] set up on four-taxon imbalanced subtrees [(((P1,P2),P3),P4)], where the P3 taxon is expected to share derived alleles with either of the two sister species, P1 or P2, with equal probability. An excess of allele sharing among P3 with either of these two taxa (that is, discordance) is interpreted as evidence of admixture[50].

We note, though, that in the context of multiple possible subtrees from a larger phylogenetic context, caution should be taken when interpreting *D*-statistics, as they can have a high false-positive rate caused by shared ancestry among potential introgressive donors[83].

We implemented *D*-statistic calculations in the ipa.baba module of ipyrad-analysis[56], which is specifically designed for analysis of RAD-seq datasets. In this tool, all loci shared among samples in each subtree are extracted and filtered from the database to apply minimum sample coverage thresholds. Site patterns matching ABBA and BABA are calculated using allele frequencies when multiple samples represent a taxon in the analysis. We required a minimum coverage of one sample per taxon per test, but most taxa were represented by multiple samples. To assess the significance, we performed 500 bootstrap replicates for each test where RAD loci are resampled with replacement. A *Z*-score is then calculated as the number of bootstrap standard deviations that *D* deviates from zero. For visualization we show a scatterplot of the 500 bootstrap replicates and coloured points darker when the *Z*-score was significant using $Z > 5$ as a cutoff (Fig. 3). Reproducible code to replicate our results and figures is available in Jupyter Notebooks: https://github.com/eaton-lab/Oreinotinus-phylogeny. Statistics for each test are available in Supplementary Table 4.

We first set up a series of 15 tests to investigate admixture between species from neighbouring regions (Fig. 3a). In these tests, P1 and P2 were selected as species from the same region that differ in leaf form. P3 was selected as a species from a neighbouring region that shares a leaf form with the P2 taxon. All tests used *V. dentatum*, the outgroup to *Oreinotinus* from eastern North America, as the P4 outgroup taxon. Across this series of tests we found only one significant result, showing admixture between *V. sulcatum* (Oaxaca) and *V. tiliaefolium* (eastern Mexico) (Fig. 3b). See below and Extended Data Fig. 7 for a further investigation of admixture in *V. sulcatum*. This series of tests included a mean of 131,024 SNPs per test (s.d. = 64,309), of which 2.99% were discordant (ABBA or BABA; s.d. = 0.4%), composing an average of 1,983 ABBA sites and 1,925 BABA sites per test.

A second series of tests was also performed on an 'alternative phylogenetic hypothesis', wherein the same four-taxon sets were analysed, but the P1, P2 and P3 subtree topology was changed to instead represent a case where two species with similar leaf form from neighbouring regions are assumed to share a most recent common ancestor (that is, set as P1 and P2; Fig. 3c). If our inferred phylogenetic hypothesis is correct then we should see an excess of ABBA sites, wherein the two species from the same region share more alleles in common. This series of tests had the same number of SNPs as above, but a greater proportion are discordant (mean = 3.89%, s.d. = 0.64%), leading to a mean *D* of 0.22 (s.d. = 0.14) and mean *Z*-score of 29.25 (s.d. = 21.94).

To test for evidence of admixture among species within the same region, we set up a series of tests sampling three taxa in each region matching the required imbalanced topology shape based on our inferred concatenation tree for *Oreinotinus* (Fig. 3e). These tests were not set up with specific regard to leaf form, so an excess of either ABBA or BABA is simply interpreted as evidence of admixture. Finally, we used *D*-statistics to investigate the source of admixture in *V. sulcatum* (Extended Data Fig. 7). A series of 13 tests was set up to compare *V. sulcatum* as the P3 taxon relative to 2 contrasted P1 and P2 samples in eastern Mexico, or 1 in eastern Mexico and 1 in western Mexico. We also tested in the reverse direction with eastern and western Mexico taxa as P3, and *V. sulcatum* as the P2 taxon contrasted against either a distant relative (*V. jucundum* from Chiapas) or close relative (*V. fuscum* from Oaxaca).

Network analyses. To estimate a phylogenetic network, we generated trees using RAxML[61] for 178 individuals along 643 windows using the module ipa.treeslider included in ipa[56] (Fig. 3c and Supplementary Fig. 10). Those windows were extracted from the 100 largest scaffolds of our entire dataset. Each window had a size of 2 Mb without overlapping. We only kept windows with at least 10 SNPs and a minimum coverage of 9 samples. We further filter the sliding windows to only keep those that have more than 300 SNPs. All samples belonging to the same species were reduced to a unique representative sample using consensus reduction. In total, 357 trees were recovered and used as inputs to reconstruct the network. For this we used NANUQ[84] incorporated in MSCquartets package[85] for R[86], setting an alpha of 1×10^{-5} and a beta of 0.95. We used SplitsTree[51] to plot the final network using default parameters.

Climate analyses for *V. jucundum*, *V. lautum* and *V. hartwegii*. We assessed the environmental niche space of three species representing three distinct leaf types on the Central Plateau region of Chiapas, Mexico: *V. hartwegii* (IGE), *V. jucundum* (LPT) and *V. lautum* (SGE) (Fig. 4 and Extended Data Fig. 8). We characterized their environmental niche in two ways. First, we assembled georeferenced localities for each species within the region by combining our own collection localities with all preserved specimen records in GBIF[87]. We then filtered collection points using the R package CoordinateCleaner[66] and mapped the remaining localities. This yielded 230 vetted localities plus 5 additional localities from our climate stations in Chiapas (see below). We extracted 19 standard bioclimatic variables for all vetted localities from CHELSA (climatologies at high resolution for the Earth's land surface areas)[88,89] at 1 km² resolution. Elevation data were obtained using the digital elevation model from the Shuttle Radar Topography Mission[90,91] at 90 m²

resolution. The raster value of each data layer at each locality point was extracted in ArcGIS Pro using an ArcPy script, and we analysed the ArcGIS output in R, using *t*-tests to compare pairwise means for each set of species and each climate variable, and plotted climate variable distributions for each species in ggplot[92]. Prior to extracting environmental values, all data points and raster layers were projected into the coordinate system Mexico IRTF 2008 UTM 15N, the coordinate system for which elevational measurements in central Mexico should experience the least distortion.

Second, to investigate local variation in environment at a finer scale we installed five portable climate stations (Meter Environment) in Chiapas at two *V. jucundum* (LPT) sites, two *V. lautum* (SGE) sites and one *V. hartwegii* (IGE) site (Fig. 4e,f). Climate stations were installed centrally within healthy, natural mono-specific *Viburnum* stands; *V. jucundum* and *V. lautum* stations were installed in May 2016, and the *V. hartwegii* station was added in January 2018. Stations continually recorded data until our last upload in February 2019. Each climate station included an air temperature sensor and a leaf wetness sensor connected to a data logger (Decagon/Meter Group ATMOS 14, PHYTOS 31, Em50 Data Logger). The temperature and leaf wetness sensors were attached to tree trunks at a similar height to the *Viburnum* canopy, around 2–3 m above ground. Temperature and leaf wetness measurements were logged hourly. We analysed station data in R. We compared temperature and leaf wetness on an hourly, daily and monthly basis between sites using *t*-tests with a significance level of 0.05 to test for significant overlap between climate measurements from each site. In addition, for each climate station area, we estimated canopy cover as the proportion of open sky present below the canopy and used this as a proxy for light environment (Fig. 4d). For each site, we identified three to five individuals of the species being monitored and took two fisheye photographs at each individual plant: one picture 1 m north of the individual and one picture 1 m to the south. Pictures were taken using a digital SLR equipped with a fisheye lens using methods detailed in ref. [93]. Images were binarized and the percentage of white pixels was used as a proxy for percent open canopy. We tested for significant differences in canopy cover between sites using R and plotted values using ggplot[92].

Climate analyses across the geographic range of *Oreinotinus*. To determine if IGE, LPT and SGE forms are generally differentiated as they are in our detailed three-species comparison in Chiapas, we compared CHELSA variables across the entire geographic ranges of *V. hartwegii*, *V. jucundum* and *V. lautum* (Supplementary Fig. 12) with IGE, LPT and SGE species across the range of *Oreinotinus* (Extended Data Figs. 9 and 10) and also within individual regions (Supplementary Figs. 13–15). We analysed the full dataset of 3,034 curated specimen occurrences and associated environmental variables (see above) from across the full range of *Oreinotinus*. We removed *V. venustum*, *V. hondurense* and *V. incarum*, as these species were not sampled in the tree. We also removed *V. membranaceum* and *V. hirsutum*, which had only three occurrence records each. Using pairwise *t*-tests implemented in R and visualized in ggplot, we first compared environmental conditions by area of endemism (Supplementary Fig. 11) then by ecomorph across all regions (Extended Data Figs. 9 and 10). After assessing which of the 19 CHELSA climate variables showed significant differences between ecomorphs, we focused on the following variables: mean annual temperature, mean diurnal range, temperature seasonality, temperature annual range, minimum temperature of the coldest month, mean annual precipitation, precipitation seasonality and elevation (Extended Data Figs. 9 and 10 and Supplementary Figs. 12–15). We selected these variables to include the standard climate niche measures included in comparable studies (mean annual temperature, mean annual precipitation, elevation), as well as all metrics that represent variation as opposed to averages (mean diurnal range, temperature seasonality, temperature annual range, minimum temperature of the coldest month, precipitation seasonality) on the view that these may be especially important in determining *Viburnum* distributions.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability
All data are available in the Zenodo data repository https://doi.org/10.5281/zenodo.5504439.

Code availability
Scripts for reproducible science are in the GitHub repository at https://github.com/eaton-lab/Oreinotinus-phylogeny.

References
1. Schluter, D. *The Ecology of Adaptive Radiation* (Oxford Univ. Press, 2000).
2. Losos, J. B. Adaptive radiation, ecological opportunity, and evolutionary determinism. *Am. Nat.* **175**, 623–639 (2010).

3. Losos, J. B. *Lizards in an Evolutionary Tree: Ecology and Adaptive Radiation of Anoles* (Univ. California Press, 2009).
4. Mahler, D. L., Ingram, T., Revell, L. J. & Losos, J. B. Exceptional convergence on the macroevolutionary landscape in island lizard radiations. *Science* **341**, 292–295 (2013).
5. Muschick, M., Indermaur, A. & Salzburger, W. Convergent evolution within an adaptive radiation of cichlid fishes. *Curr. Biol.* **22**, 2362–2368 (2012).
6. Wagner, C. E., Harmon, L. J. & Seehausen, O. Ecological opportunity and sexual selection together predict adaptive radiation. *Nature* **487**, 366–369 (2012).
7. Gillespie, R. Community assembly through adaptive radiation in Hawaiian spiders. *Science* **303**, 356–359 (2004).
8. Gillespie, R. G. et al. Repeated diversification of ecomorphs in Hawaiian stick spiders. *Curr. Biol.* **28**, 941–947 (2018).
9. Abrahamczyk, S. & Renner, S. S. The temporal build-up of hummingbird/plant mutualisms in North America and temperate South America. *BMC Evol. Biol.* **15**, 104 (2015).
10. Sinnott-Armstrong, M. A. et al. Fruit syndromes in *Viburnum*: correlated evolution of color, nutritional content, and morphology in bird-dispersed fleshy fruits. *BMC Evol. Biol.* **20**, 7 (2020).
11. Edwards, E. J. Evolutionary trajectories, accessibility and other metaphors: the case of C₄ and CAM photosynthesis. *New Phytol.* **223**, 1742–1755 (2019).
12. Hughes, C. & Eastwood, R. Island radiation on a continental scale: exceptional rates of plant diversification after uplift of the Andes. *Proc. Natl Acad. Sci. USA* **103**, 10334–10339 (2006).
13. Givnish, T. J. et al. Adaptive radiation, correlated and contingent evolution, and net species diversification in Bromeliaceae. *Mol. Phylogenet. Evol.* **71**, 55–78 (2014).
14. Lagomarsino, L. P. et al. The abiotic and biotic drivers of rapid diversification in Andean bellflowers (Campanulaceae). *New Phytol.* **210**, 1430–1442 (2016).
15. Schenk, J. J. The next generation of adaptive radiation studies in plants. *Int. J. Plant Sci.* **182**, 245–262 (2021).
16. Givnish, T. J. et al. Origin, adaptive radiation and diversification of the Hawaiian lobeliads (Asterales: Campanulaceae). *Proc. R. Soc. B* **276**, 407–416 (2009).
17. Drummond, C. S., Eastwood, R. J., Miotto, S. T. & Hughes, C. E. Multiple continental radiations and correlates of diversification in *Lupinus* (Leguminosae): testing for key innovation with incomplete taxon sampling. *Syst. Biol.* **61**, 443–460 (2012).
18. Roquet, C. et al. Replicated radiations of the alpine genus *Androsace* (Primulaceae) driven by range expansion and convergent key innovations. *J. Biogeogr.* **40**, 1874–1886 (2013).
19. DiVittorio, C. T. et al. Natural selection maintains species despite frequent hybridization in the desert shrub *Encelia*. *Proc. Natl Acad. Sci. USA* **117**, 33373–33383 (2020).
20. Knotek, A. et al. Parallel alpine differentiation in *Arabidopsis arenosa*. *Front. Plant Sci.* **11**, 561526 (2020).
21. Pouchon, C. et al. Phylogenetic signatures of ecological divergence and leapfrog adaptive radiation in *Espeletia*. *Am. J. Bot.* **108**, 113–128 (2021).
22. Donoghue, M. J. *Systematic Studies in the Genus Viburnum*. PhD dissertation, Harvard Univ. (1982).
23. Donoghue, M. J., Bell, C. D. & Winkworth, R. C. The evolution of reproductive characters in Dipsacales. *Int. J. Plant Sci.* **164**, S453–S464 (2003).
24. Landis, M. J. et al. Joint estimation of geographic movements and biome shifts during the global diversification of *Viburnum*. *Syst. Biol.* **70**, 67–85 (2021).
25. Clement, W. L. et al. A chloroplast tree for *Viburnum* (Adoxaceae) and its implications for phylogenetic classification and character evolution. *Am. J. Bot.* **101**, 1029–1049 (2014).
26. Spriggs, E. L. et al. Temperate radiations and dying embers of a tropical past: evidence from *Viburnum* diversification. *New Phytol.* **207**, 340–354 (2015).
27. Moeglein, M. et al. Evolutionary dynamics of genome size in a radiation of woody plants. *Am. J. Bot.* **107**, 1527–1541 (2020).
28. Donoghue, M. J. & Sanderson, M. J. Confluence, synnovation, and depauperons in plant diversification. *New Phytol.* **207**, 260–274 (2015).
29. Nurk, N. M. et al. Diversification in evolutionary arenas – assessment and synthesis. *Ecol. Evol.* **10**, 6163–6182 (2020).
30. Karger, D. D. et al. Limited protection and ongoing loss of cloud forest biodiversity and ecosystems worldwide. *Nat. Ecol. Evol.* **5**, 854–862 (2021).
31. Mastretta-Yanes, A. et al. Biodiversity in the Mexican highlands and the interaction of geology, geography and climate within the trans-Mexican volcanic belt. *J. Biogeogr.* **42**, 1586–1600 (2015).
32. Edwards, E. J. et al. Convergence, consilience, and the evolution of temperate deciduous forests. *Am. Nat.* **190**, S87–S104 (2017).
33. Ree, R. H. et al. A likelihood framework for inferring the evolution of geographic range on phylogenetic trees. *Evolution* **59**, 2299–2311 (2005).
34. Weber, M. G., Donoghue, M. J., Clement, W. L. & Agrawal, A. A. Phylogenetic and experimental tests of interactions among mutualistic plant defense traits in *Viburnum* (Adoxaceae). *Am. Nat.* **180**, 450–463 (2012).
35. Parkhurst, D. F. & Loucks, O. L. Optimal leaf size in relation to environment. *J. Ecol.* **60**, 505–537 (1972).
36. Givnish, T. J. in *Topics in Plant Population Biology* (eds Solbrig, O. T. et al.) 375–407 (Columbia Univ. Press, 1979).
37. Givnish, T. J. Comparative studies of leaf form: assessing the relative roles of selective pressures and phylogenetic constraints. *New Phytol.* **106**, 131–160 (1987).
38. Wright, I. J. et al. Global climatic drivers of leaf size. *Science* **357**, 917–921 (2017).
39. Levin, D. A. The role of trichomes in plant defense. *Q. Rev. Biol.* **48**, 3–15 (1973).
40. Ehleringer, J. in *Biology and Chemistry of Plant Trichomes* (eds Rodriguez, E. et al.) 113–132 (Plenum Press, 1984).
41. Brewer, C. A., Smith, W. K. & Vogelmann, T. C. Functional interaction between leaf trichomes, leaf wettability and the optical properties of water droplets. *Plant Cell Environ.* **14**, 955–962 (1991).
42. Bickford, C. P. Ecophysiology of leaf trichomes. *Funct. Plant Biol.* **43**, 807–814 (2016).
43. Bailey, I. W. & Sinnott, E. W. The climatic distribution of certain types of angiosperm leaves. *Am. J. Bot.* **3**, 24–39 (1916).
44. Royer, D. L. & Wilf, P. Why do toothed leaves correlate with cold climates? Gas exchange at leaf margins provides new insights into a classic paleotemperature proxy. *Int. J. Plant Sci.* **167**, 11–18 (2006).
45. Edwards et al. Unpacking a century-old mystery: winter buds and the latitudinal gradient in leaf form. *Am. J. Bot.* **103**, 975–978 (2016).
46. Stayton, C. T. The definition, recognition, and interpretation of convergent evolution, and two new measures for quantifying and assessing the significance of convergence. *Evolution* **69**, 2140–2153 (2015).
47. Lamichhaney, S. et al. Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature* **518**, 371–375 (2015).
48. Malinsky, M. et al. Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nat. Ecol. Evol.* **2**, 1940–1955 (2018).
49. Edelman, N. B. et al. Genomic architecture and introgression shape a butterfly radiation. *Science* **366**, 594–599 (2019).
50. Durand, E. Y. et al. Testing for ancient admixture between closely related populations. *Mol. Biol. Evol.* **28**, 2239–2252 (2011).
51. Huson, D. H. & Bryant, D. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23**, 254–267 (2006).
52. Clement, W. L. et al. Parallelism in endocarp form sheds light on fruit syndrome evolution in *Viburnum*. *Syst. Bot.* **46**, 504–517 (2021).
53. Donoghue, M. J. Flowering times in *Viburnum*. *Arnoldia* **40**, 2–22 (1980).
54. Spriggs, E. L. et al. Differences in flowering time maintain species boundaries in a continental radiation of *Viburnum*. *Am. J. Bot.* **106**, 833–849 (2019).
55. Rundell, R. J. & Price, T. D. Adaptive radiation, non-adaptive radiation, ecological speciation and non-ecological speciation. *Trends Ecol. Evol.* **24**, 394–399 (2009).
56. Eaton, D. A. R. & Overcast, I. ipyrad: interactive assembly and analysis of RADseq datasets. *Bioinformatics* **36**, 2592–2594 (2020).
57. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
58. Goltsman, E., Ho, I. Y. & Rokhsar, D. S. Meraculous-2D: haplotype-sensitive assembly of highly heterozygous genomes. Preprint at https://arxiv.org/abs/1703.09852 (2017).
59. Ranallo-Benavidez, T. R., Jaron, K. S. & Schatz, M. C. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat. Commun.* **11**, 1432 (2020).
60. Putnam, N. H. et al. Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome Res.* **26**, 342–350 (2016).
61. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
62. Eaton, D. A. R. Toytree: a minimalist tree visualization and manipulation library for Python. *Methods Ecol. Evol.* **11**, 187–191 (2020).
63. Zhang, C., Rabiee, M., Sayyari, E. & Mirarab, S. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinform.* **19**, 153 (2018).
64. Tamura, T., Tao, Q. & Kumar, S. Theoretical foundation of the RelTime method for estimating divergence times from variable evolutionary rates. *Mol. Biol. Evol.* **35**, 1770–1782 (2018).
65. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).
66. Zizka, A. et al. CoordinateCleaner: standardized cleaning of occurrence records from biological collection databases. *Methods Ecol. Evol.* **10**, 744–751 (2019).
67. Edler, D., Guedes, T., Zizka, A., Rosvall, M. & Antonelli, A. Infomap Bioregions: interactive mapping of biogeographical regions from species distributions. *Syst. Biol.* **66**, 197–204 (2017).
68. Höhna, S. et al. RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Syst. Biol.* **65**, 726–736 (2016).

69. Landis, M. J., Freyman, W. A. & Baldwin, B. G. Retracing the Hawaiian silversword radiation despite phylogenetic, biogeographic, and paleogeographic uncertainty. *Evolution* **72**, 2343–2359 (2018).

70. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for dimension reduction. Preprint at http://arxiv.org/abs/1802.03426 (2020).

71. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).

72. Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

73. Spriggs, E. L., Schmerler, S. B., Edwards, E. J. & Donoghue, M. J. Leaf form evolution in *Viburnum* parallels variation within individual plants. *Am. Nat.* **191**, 235–249 (2018).

74. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).

75. Maloof, J. N., Nozue, K., Mumbach, M. R. & Palmer, C. M. LeafJ: an ImageJ plugin for semi-automated leaf shape measurement. *J. Vis. Exp.* (71), e50028, https://doi.org/10.3791/50028 (2013).

76. Lever, J., Krzywinski, M. & Altman, N. Principal component analysis. *Nat. Methods* **14**, 641–642 (2017).

77. Ingram, T. & Mahler, D. L. SURFACE: detecting convergent evolution from comparative data by fitting Ornstein-Uhlenbeck models with stepwise Akaike information criterion. *Methods Ecol. Evol.* **4**, 416–425 (2013).

78. Kassambra, A. ggpubr: 'ggplot2' Based Publication Ready Plots https://cran.r-project.org/web/packages/ggpubr/index.html (2020).

79. Revell, L. J. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223 (2012).

80. Louca, S. & Doebeli, M. Efficient comparative phylogenetics on large trees. *Bioinformatics* **34**, 1053–1055 (2018).

81. Stayton, C. T. Package 'convevol': Analysis of Convergent Evolution. Version 1.3 https://mirror.linux.duke.edu/cran/web/packages/convevol/convevol.pdf (2018).

82. Brown, J. M. & Thomson, R. C. Evaluating model performance in evolutionary biology. *Ann. Rev. Ecol. Evol. Syst.* **49**, 95–114 (2018).

83. Eaton, D. A. R., Hipp, A. L., González-Rodríguez, A. & Cavender-Bares, J. Historical introgression among the American live oaks and the comparative nature of tests for introgression. *Evolution* **69**, 2587–2601 (2015).

84. Allman, E. S., Baños, H. & Rhodes, J. A. NANUQ: a method for inferring species networks from gene trees under the coalescent model. *Algorithms Mol. Biol.* **14**, 24 (2019).

85. Rhodes, J. A., Baños, H., Mitchell, J. D. & Allman, E. S. MSCquartets 1.0: quartet methods for species trees and networks under the multispecies coalescent model in R. *Bioinformatics* https://doi.org/10.1093/bioinformatics/btaa868 (2020).

86. R Core Team *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2017).

87. *GBIF Occurrence Download* https://doi.org/10.15468/dl.prz2j3 (GBIF.org, 2021).

88. Karger, D. N. et al. Climatologies at high resolution for the Earth land surface areas. *Sci. Data* **4**, 170122 (2017).

89. Karger D. N. et al. Data from: Climatologies at high resolution for the Earth land surface areas. *Dryad* https://doi.org/10.5061/dryad.kd1d4 (2018).

90. Farr, T. G. et al. The shuttle radar topography mission. *Rev. Geophys.* **45**, RG2004 (2007).

91. *Shuttle Radar Topography Mission (SRTM) Global* (Open Topography, 2013); https://portal.opentopography.org/datasetMetadata?otCollectionID=OT.042013.4326.1

92. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* 2nd edn (Springer, 2009).

93. Glatthorn, J. & Beckschäfer, P. Standardizing the protocol for hemispherical photographs: accuracy assessment of binarization algorithms. *PLoS ONE* **9**, e111924 (2014).

94. Wilke, C. O. Package 'ggridges': Ridgeline Plots in 'ggplot2'. Version 0.5.3 https://CRAN.R-project.org/package=ggridges (2021).

Author contributions

Study design, funding, field studies, analyses and writing were carried out by M.J.D., D.A.R.E. and E.J.E. Additional contributions and analyses were performed by C.A.M.-L., M.J.L., P.W.S., J.R.G., N.M.H. and M.K.M. Field studies were also carried out by P.W.S., M.E.O., N.I.C., M.K.M., M.C., A.S.R. and W.L.C. Additional editing was carried out by C.A.M.-L., M.J.L., M.E.O. and N.M.H.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41559-022-01823-x.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41559-022-01823-x.

Correspondence and requests for materials should be addressed to Michael J. Donoghue, Deren A. R. Eaton or Erika J. Edwards.
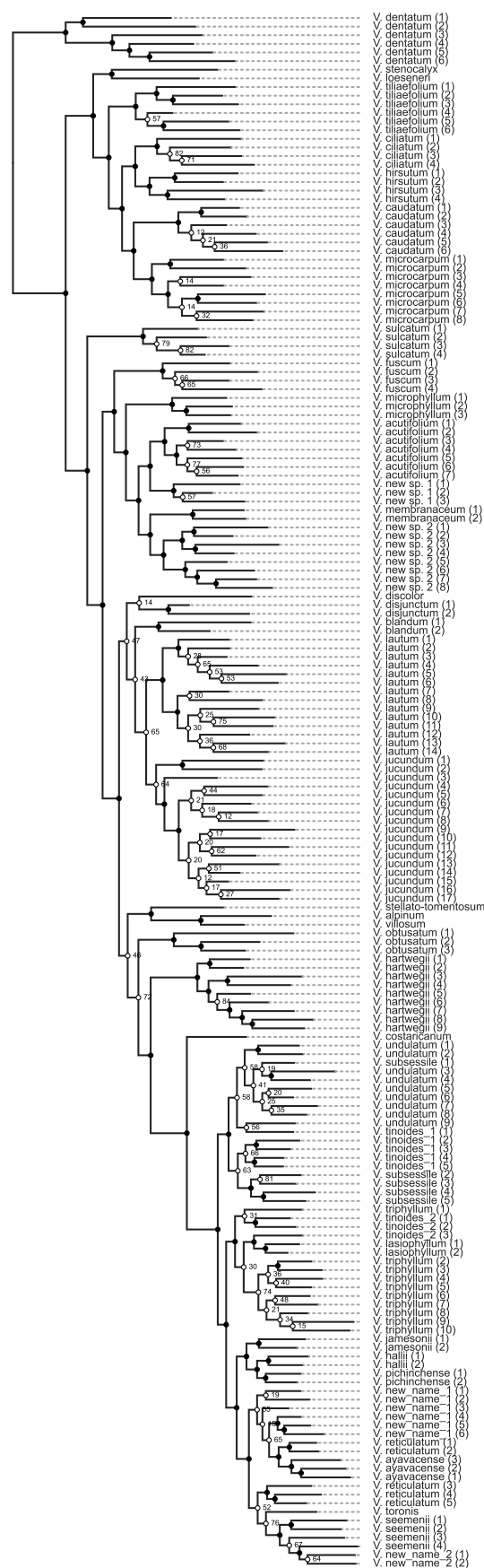
Peer review information *Nature Ecology & Evolution* thanks Julia Day and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
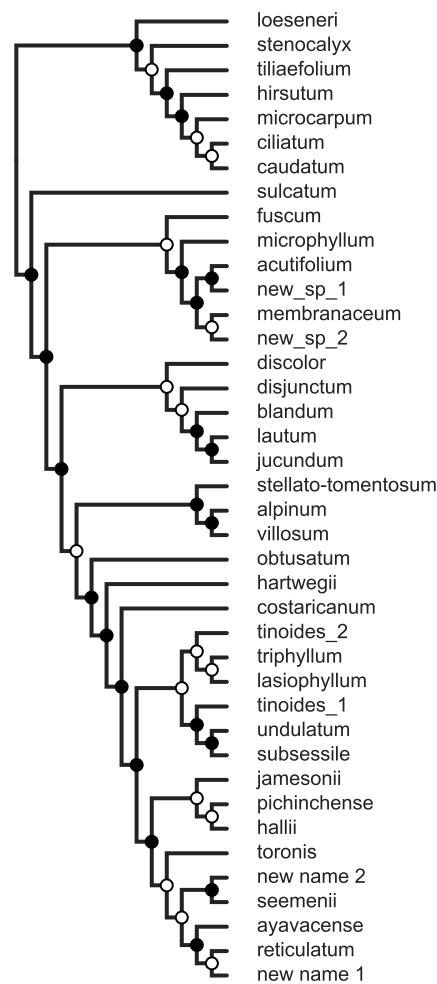
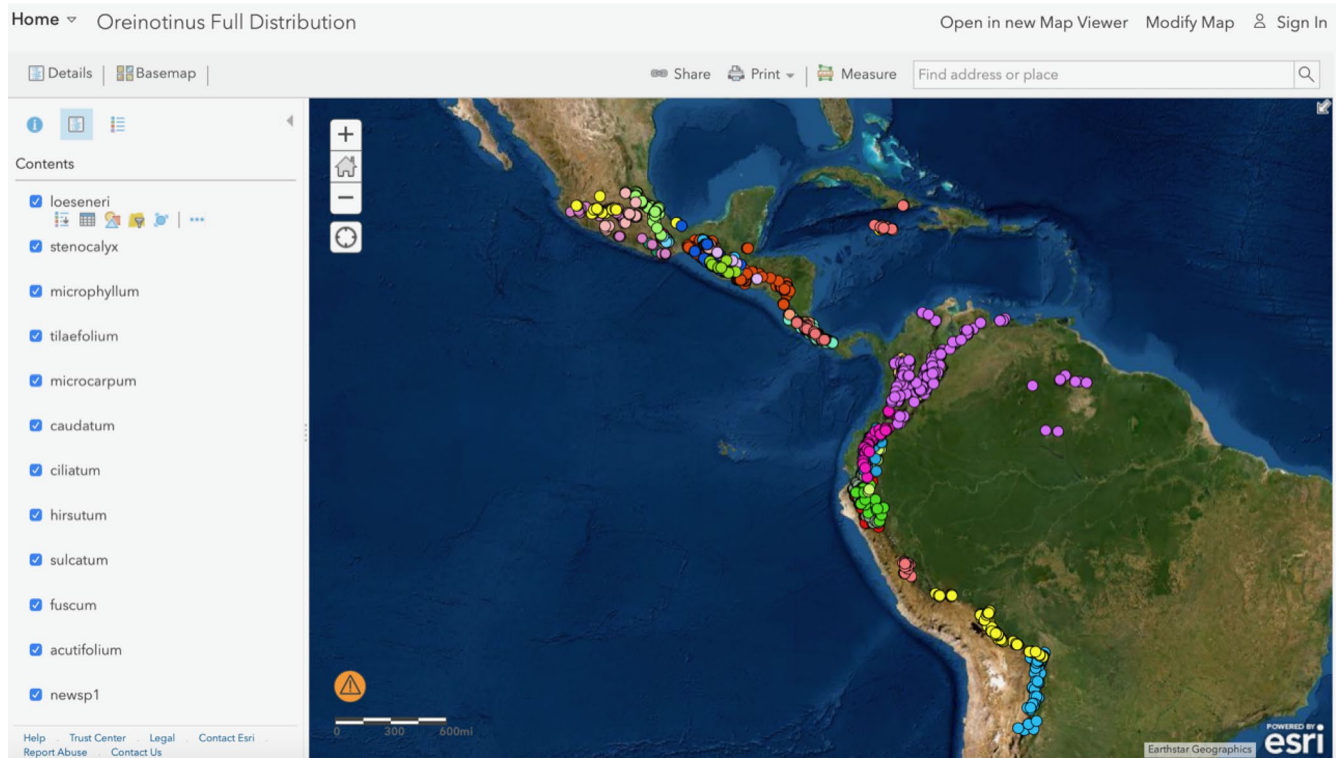**Extended Data Fig. 1 | See next page for caption.**

Extended Data Fig. 1 | Individual-level RAD-seq tree for *Oreinotinus*. ML tree inferred from concatenated RAD-seq data from 180 individuals of the *Oreinotinus* clade plus outgroups (see Methods). Specimen numbers are linked to voucher specimen information in Supplement Table 3. Bootstrap support values are shown at the nodes; black dots mark support of >85%. Accessions from multiple populations of the same morphology-based species form generally well-supported clades in all but five instances; these cases are discussed in the main text and in in Supplement Note 1.

Extended Data Fig. 2 | Astral species tree for *Oreinotinus*. Multispecies coalescent tree of 356 trees with more than 300 SNPs produced by ASTRAL-III (see Methods). Black nodes are supported by a local posterior probability > =85%.
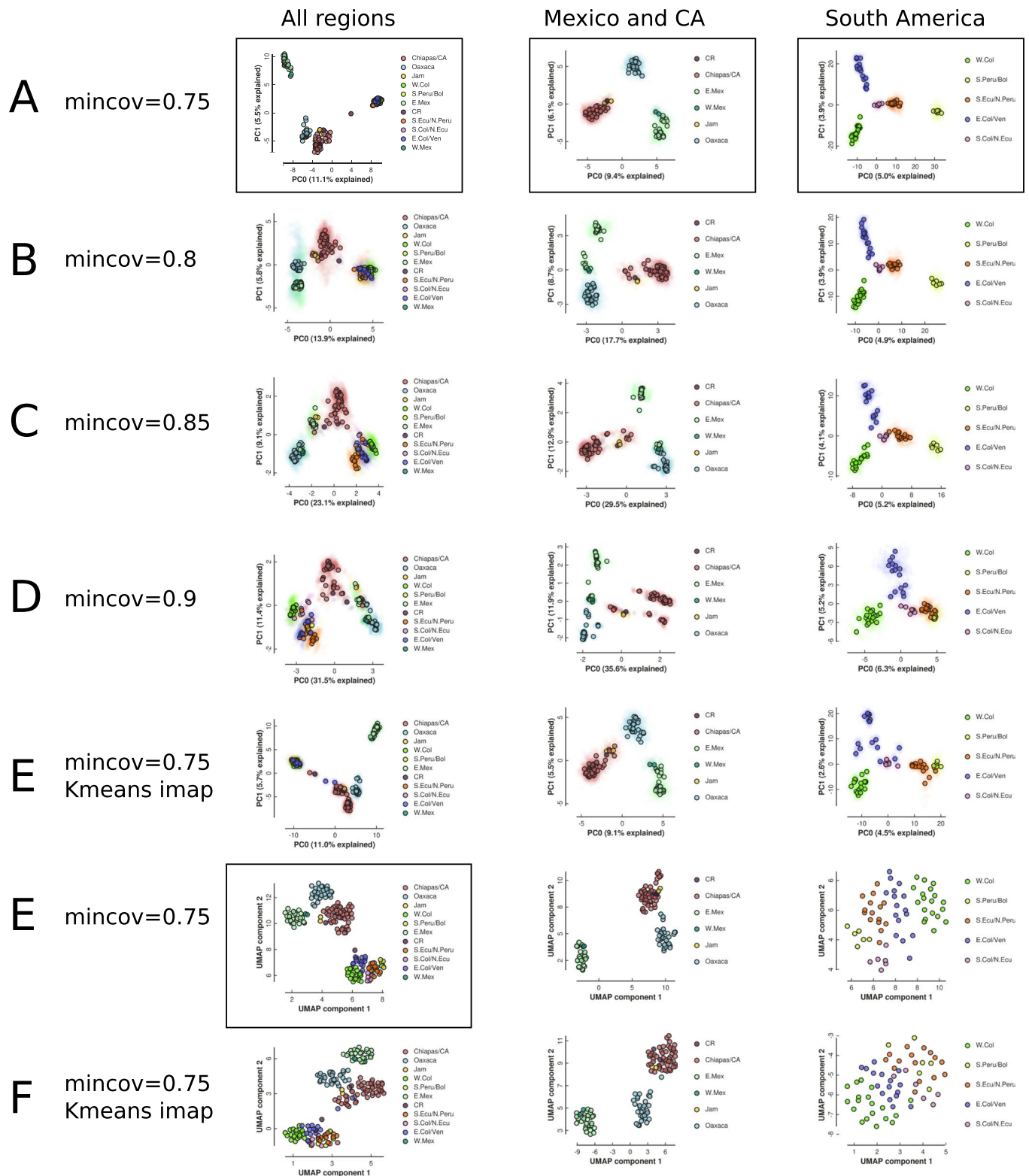
To access the map please copy and paste this URL into your browser:

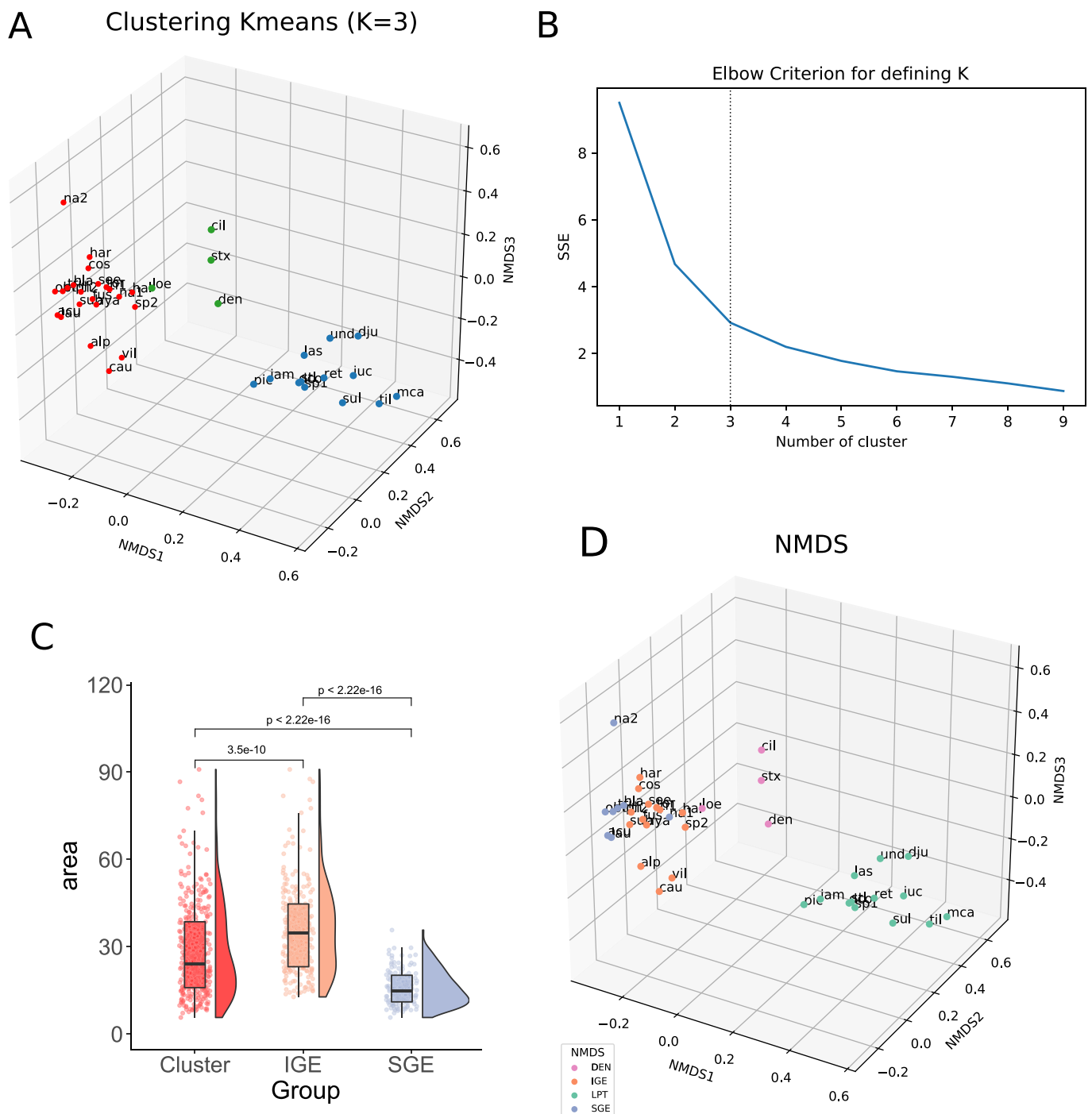# https://arcg.is/1OLDXD0

Extended Data Fig. 3 | Interactive species range map. Points on the map represent all *Oreinotinus* localities included in our final vetted dataset, coloured by species and mapped using ArcGIS Pro (see Methods). Species are ordered by region in the legend, from north to south. Individual species ranges can be viewed by unchecking the boxes located in the left-hand panel to temporarily remove co-occurring species from the display. Clicking on a point will open a pop-up menu with metadata on the specimen attached to that point, including latitude and longitude, year of collection, institution and, where available, collector and specimen number. To view the distributions against a different background layer, such as topography or political boundaries, click 'Basemap' in the top bar.
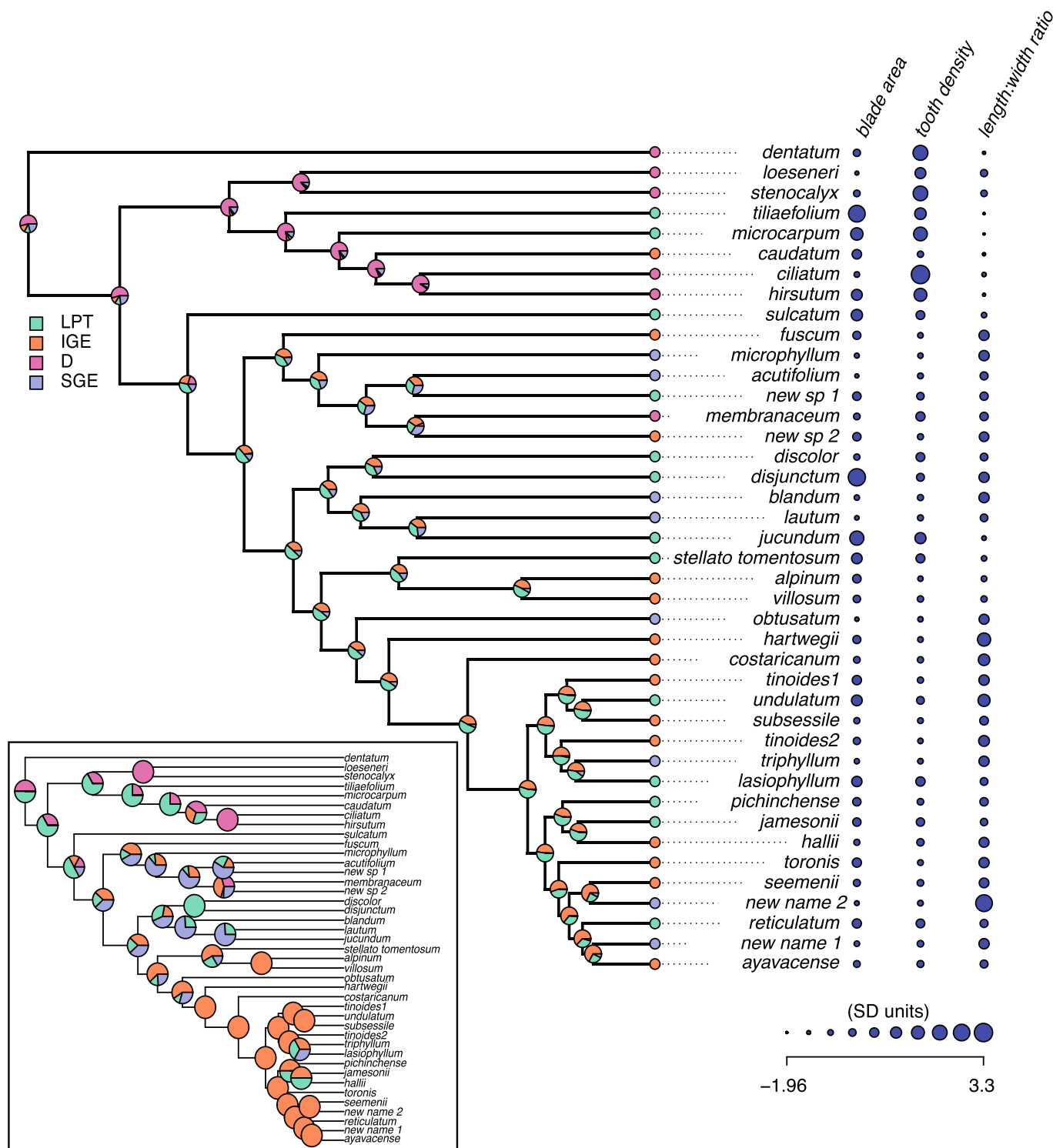
Extended Data Fig. 4 | Dimensionality reduction methods (UMAP and PCA) applied to SNP datasets with different filtering and imputation.
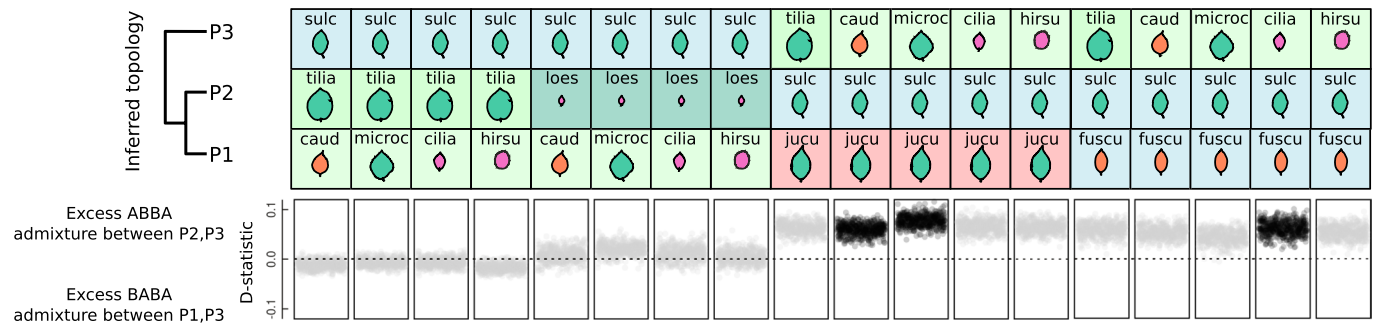The subpanels included in Fig. 1 are outlined with black boxes (see Methods). (A) Our primary analyses filtered SNPs to require > =75% sample coverage at all SNPs and imputed the remaining data. Filtering with a higher minimum coverage thresholds (B-D) yielded highly similar results but began to lose statistical power when few SNPs were available at the highest thresholds. (E) An alternative K-means clustering method for assigning samples to groups for imputation recovered highly similar results as when using *a priori* population assignments. (E) UMAP projections show both local and global structure in the 'All regions' dataset, but do not reveal additional structure in the smaller datasets. These analyses also yielded similar results when imputed with Kmeans clustering (F).

Extended Data Fig. 5 | Differentiation of leaf ecomorphs in *Oreinotinus* using nonmetric multidimensional scaling (NMDS). A and C are based on three quantitative variables [leaf blade area, aspect ratio (leaf length/width) and the density of marginal teeth] and two categorical variables (a four-state character reflecting pubescence density and a two-state character for trichome type (see Methods). (A) Three clusters are identified by NMDS with K = 3; note that this lumps our IGE and SGE categories. (B) K = 3 is chosen as optimal by the inflection point in the inertia (sum of distances of samples to their closest cluster center); this represents a cutoff point where the inertia is not improved by adding extra Ks. (C) Within the IGE-SGE cluster, IGE and SGE species differ significantly in leaf blade area; these two ecomorphs are recognized on this basis for further phylogenetic and climatic comparisons. (D) Colours mark the four leaf ecomorphs recognized in this study (DEN, LPT, IGE, SGE) based on the totality of evidence; see main text and Fig. 2).
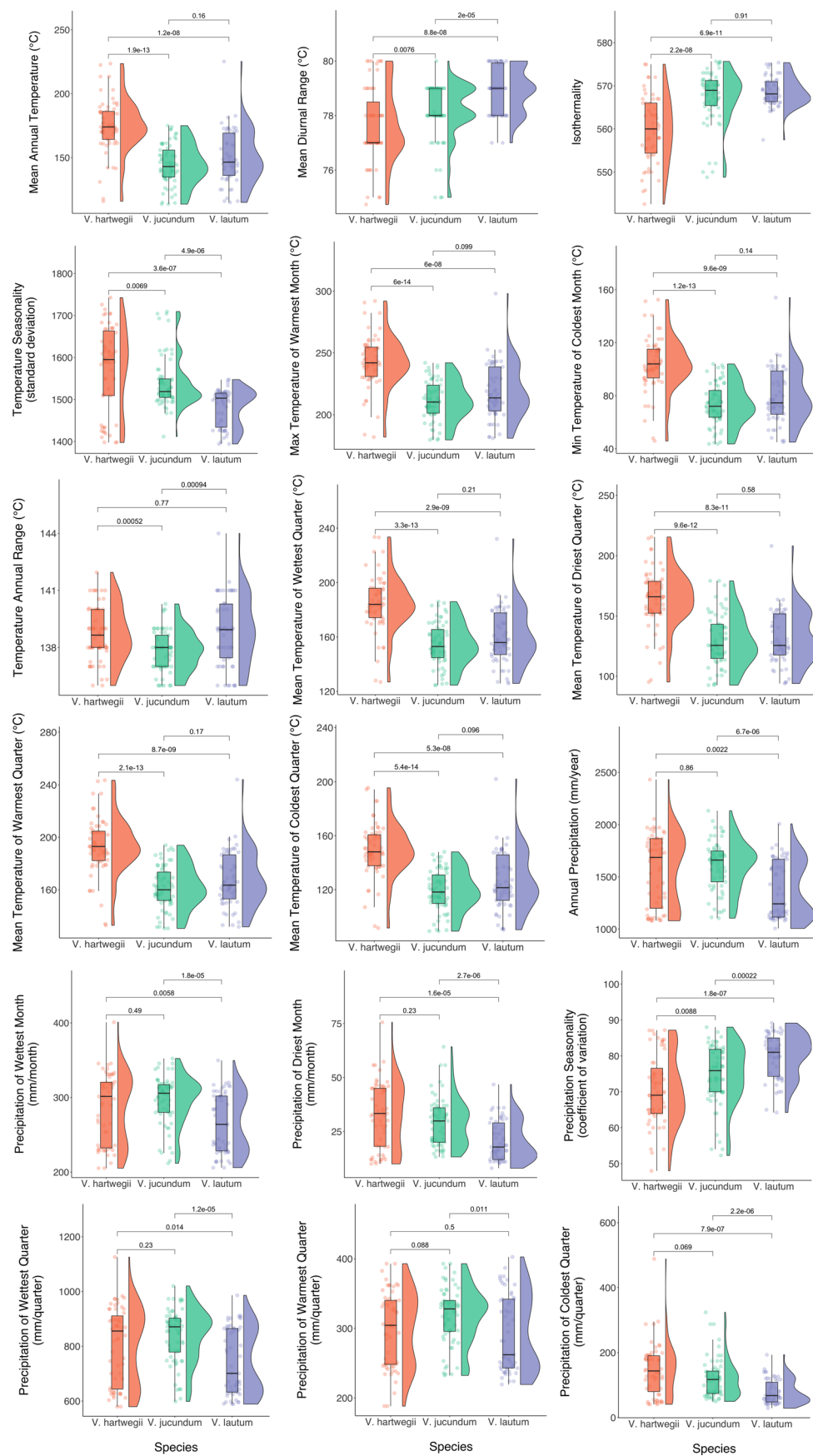
Extended Data Fig. 6 | Ancestral reconstruction of leaf ecomorphs. Ancestral states shown on the main tree represent 1,000 stochastic character mappings using ML transition matrix and the time-calibrated *Oreinotinus* phylogeny (Fig. 1). The smaller inset tree shows ancestral state reconstructions using parsimony on the same phylogeny. Our three quantitative characters are highlighted along the tips, with trait values standardized to a mean of zero and variance of 1. Bayesian stochastic mappings that account for phylogenetic uncertainty estimated a posterior median (and HPD95% interval) of 59 (28, 114) leaf form transitions (Supplement Note 3).
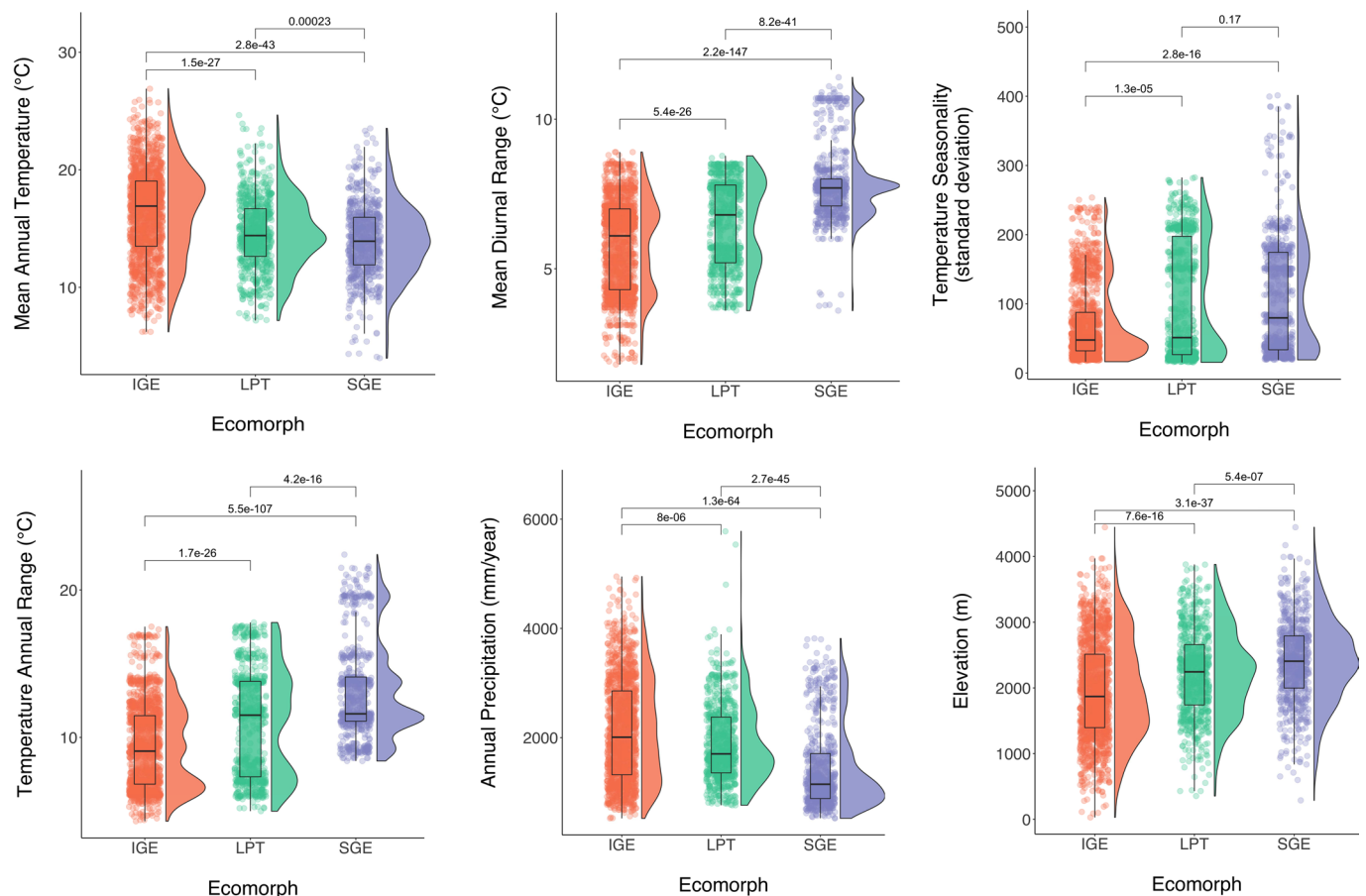
Extended Data Fig. 7 | Additional ABBA-BABA tests focused on *V. sulcatum*. Tests 1-8 show little discordance, suggesting that *V. sulcatum* has not been an introgressive donor into Eastern Mexican species of *Viburnum*. By contrast, tests 9-18 show much greater discordance when Eastern Mexican species are tested as introgressive donors into *V. sulcatum* compared to a taxon from Chiapas (*V. jucundum*) or Oaxaca (*V. acutifolium*). This suggests that *V. sulcatum* has introgressed with one or more species from the Eastern Mexican clade, but significance at the Z > 5 level is only seen in some tests and varies depending on the P3 and P1 taxa selected.
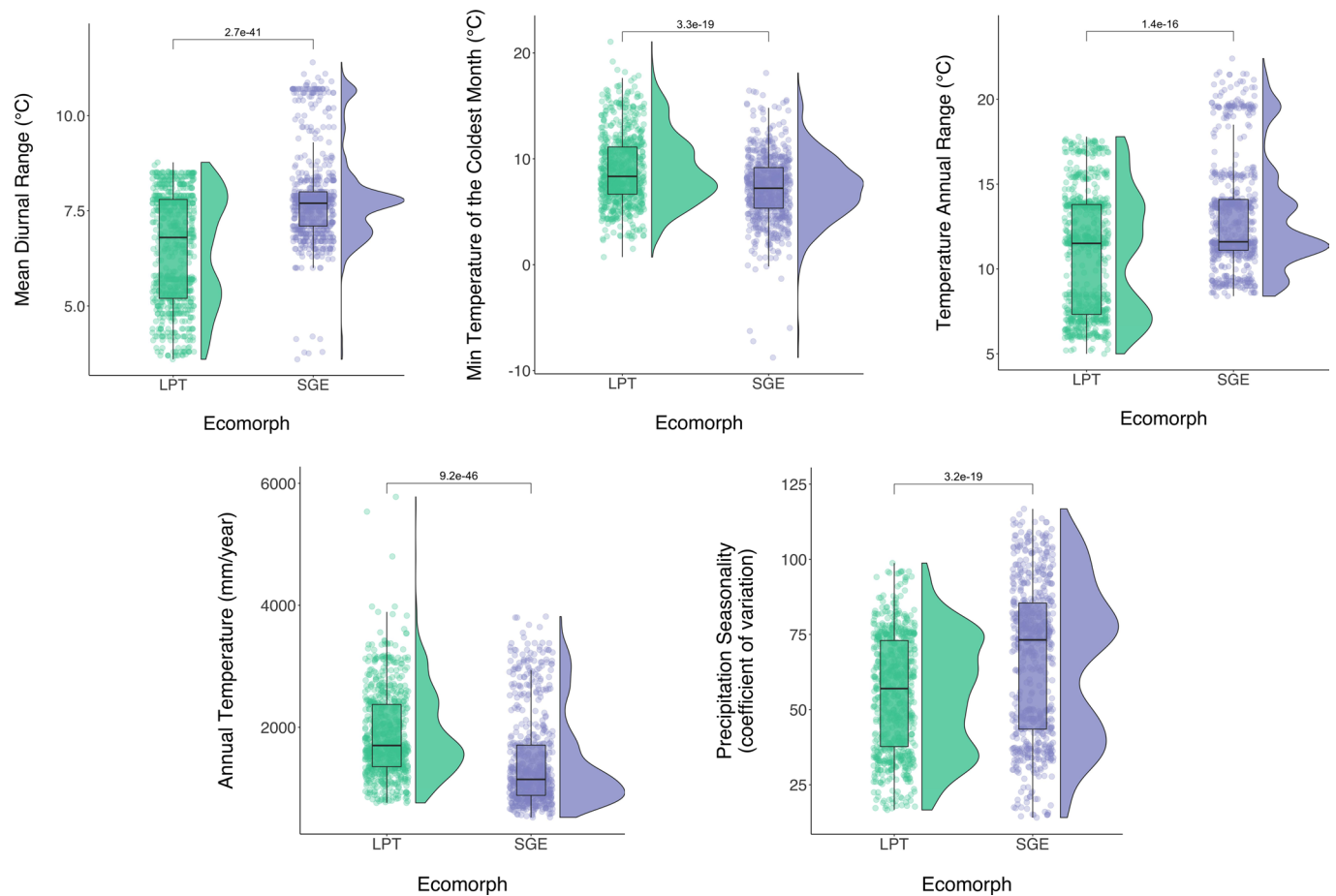
Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | CHELSA climate variables generated from geo-referenced collection localities of *V. hartwegii*, *V. jucundum* and *V. lautum* on the Central Plateau of Chiapas, Mexico (see Methods). We used the Wilcoxon rank sum test in R to compare the pairwise means of the three species. Box plot elements as follows: the center line corresponds to the median; the box limits represent upper and lower quartiles; and the whiskers span 1.5 x the interquartile range. Density ridgeline plots were drawn using the geom_density_ridges function from the ggridges R package96. N = 181 biologically independent samples (59 *V. hartwegii*, 65 *V. jucundum*, 57 *V. lautum* specimens).

Extended Data Fig. 9 | Comparison of CHELSA climate variables that differ significantly in comparing *V. hartwegii* (IGE) to *V. jucundum* (LPT) and *V. lautum* (SGE) across the geographic range of *Oreinotinus*. Compared across the entire range of *Oreinotinus*, IGE species differ significantly from LPT and SGE species in ways that align with our findings for the niches of the *V. hartwegii* in comparison with *V. jucundum* and *V. lautum* (see Methods). Specifically, we find that IGE species grow at lower elevations with higher temperatures and higher precipitation throughout the year, and lower temperature ranges and seasonality. *Viburnum hartwegii* was compared in detail with *V. jucundum* and *V. lautum* on the Central Plateau of Chiapas, Mexico (see Fig. 4). In making comparisons across the range of *Oreinotinus* we used data for the entire geographic ranges of our three focal species (that is, extending beyond the central Plateau of Chiapas; see Extended Data Fig. 3). Hypothesis testing method, box plot elements and density ridges formulation as in Extended Data Fig. 8. N = 2783 (1416 IGE, 713 LPT, 654 SGE specimens).

Extended Data Fig. 10 | Comparison of CHELSA climate variables that differ significantly between *V. jucundum* (LPT) and *V. lautum* (SGE) across the geographic range of *Oreinotinus*. Compared across the entire geographic range of *Oreinotinus*, LPT and SGE species differ significantly from one another in ways that align with our more detailed studies *V. jucundum* and *V. lautum*. Specifically, SGE species differ significantly from LPT species in their tendency to experience lower precipitation throughout the year, higher precipitation seasonality and higher seasonal and diurnal temperature ranges; SGE species may also tend to experience lower minimum temperatures in the coldest month. *Viburnum jucundum* and *V. lautum* were compared in detail on the Central Plateau of Chiapas, Mexico (see Fig. 4). In making comparisons across the range of *Oreinotinus*, we used data for the entire geographic ranges of these focal species (that is, extending beyond the central Plateau of Chiapas; see Extended Data Fig. 3. Hypothesis testing method, box plot elements and density ridges formulation as in Extended Data Fig. 8. N = 1367 (713 LPT and 654 IGE specimens).

nature portfolio

Corresponding author(s): Michael Donoghue, Deren Eaton, Erika Edwards

Last updated by author(s): 19 October 2021

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	The only software used to collect data for this study was ImageJ (v1.52 and v1.53).
Data analysis	We utilized propietary software, open source software, and custom code for data analysis. All custom scripts are available on github: https://github.com/eaton-lab/Oreinotinus-phylogeny. Software used include R, RevBayes, iPyrad, ASTRAL, RAxML, Meraculous, GenomeScope, SNAP, ipa, MEGA X, ArcGIS, and Infomap Bioregions, as carefully detailed in our supplemental methods.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All data are available in the Zenodo data repository https://www.doi.org/10.5281/zenodo.5504439, and scripts for reproducible science are at https://github.com/eaton-lab/Oreinotinus-phylogeny.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	This is a phylogenetic study. Wild plants were collected in the field and brought back to the laboratory for DNA extraction and sequencing. Herbarium vouchers were made. Leaf data were collected from images of herbarium specimens and photographs of plants in the field.
Research sample	A "research sample" would be a collection of a fertile branch of an individual, with an herbarium specimen made from the branch and used for morphological analysis, and additional leaves dried in silica for later DNA extraction and sequencing.
Sampling strategy	This was not an experiment- our sampling strategy was to locate as many distinct populations of Oreinotinus species as we could, and collect from all of them. For the leaf form analyses, we aimed to include one leaf each from a minimum of ten individuals per species, which was not possible in several very rare species with fewer collections.
Data collection	Specimen collection was documented in a standard way - collector name, date, location, description of collection and plant community it was found in.
Timing and spatial scale	Data collection occurred over many years and multiple collecting trips, spanning from Mexico to Bolivia. Most collections were made between 2015 and 2020.
Data exclusions	Specimens with low levels of successful sequencing were excluded from analyses, as well as some individuals that appeared to be hybrids between two species.
Reproducibility	There are no experiments in this study. Our analyses should be reproducible because our scripts and datasets have been made available.
Randomization	This was not an experiment and we didnt include radomization in any analyses.
Blinding	We followed a classical "taxonomic collections" mode in the field, and collected specimens from every population that we were able to locate. There is no "blinding" in this kind of work- we simply collect everything that we find.

Did the study involve field work? ☒ Yes ☐ No

Field work, collection and transport

Field conditions	Field conditions clearly varied as our work was done over very many years and across many different sites.
Location	Locations of all specimens collected are detailed in supplemental table 3, furthermore, specimen localities are provided to readers as an interactive map: https://github.com/eaton-lab/Oreinotinus-phylogeny/blob/main/Supplementary_materials/supp_fig_4_interactive_map.pdf
Access & import/export	All dried plant collections included in our analyses were collected and transported in compliance with national and international regulations. Specimens were imported into the US under our USDA 588 permit number PCIP-16-00455. We obtained all necessary collecting/exporting permits from the various countries we worked in (Ecuador: 2009: 001-IC-FAU/FLO-DZ3-09, 001-2009-IC-FLO-DPAP/MA, 03-09-IC-FLO-DNB/MA, 010-IC-FLO-DBAP-VS-DRLZCH-MA; Bolivia:2017: R.A. VMABCC 026/09; Jamaica: 2018: 18/27, 18/70; Peru: 2019: 246-2019-MIAGRI-SERFOR-DGGSPFFS, 003546-SERFOR). In Colombia and Mexico, collections were made under internal permits obtained by our resident collaborators.
Disturbance	There was no disturbance caused by this study. our collecting does no harm to the plants.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒ ☐	Antibodies
☒ ☐	Eukaryotic cell lines
☒ ☐	Palaeontology and archaeology
☒ ☐	Animals and other organisms
☒ ☐	Human research participants
☒ ☐	Clinical data
☒ ☐	Dual use research of concern

Methods

n/a	Involved in the study
☒ ☐	ChIP-seq
☒ ☐	Flow cytometry
☒ ☐	MRI-based neuroimaging